The perception of relevant surface cues shapes language development : evidence from typical and atypical populations

Caterina Marino

UNIVERSITÉ PARIS DESCARTES

Ecole doctorale 261: Cognition, Comportements, Conduites Humaines

Laboratoire Psychologie de la Perception

The perception of relevant surface cues shapes language development: evidence from typical and atypical populations

Thèse de Doctorat en Sciences Cognitives

Caterina Marino

Date de soutenance: 27 Novembre 2018

Membres de Jury:

Judit Gervain	CNRS Université Paris Descartes	Directrice de thèse
LouAnn Gerken	University of Arizona	Rapporteur
Julie Franck	Université de Genève	Rapporteur
Anne Christophe	CNRS École Normale Supérieure	Examinateur
Jean-Remy Hochmann	Université Claude Bernard, Lyon 1	Examinateur

Résumé

L’objectif général de cette thèse est d’explorer comment les enfantes perçoivent des traits de surface du langage à différents niveaux. Plus spécifiquement, on a examiné dans une population typique et atypique 1) la perception des traits acoustiques de bas-niveau et sa relation avec le développement du vocabulaire, et 2) la sensibilité à la fréquence des mots pour générer la représentation linguistique abstraite des catégories lexicales et de leur ordre relatif.

Selon les modèles de *bootstrapping,* il est possible d’extraire des informations structurelles et abstraites de la langue à partir de traits acoustiques disponibles dans l’input et corrélés avec la structure linguistique sous-jacente (p. ex. grammaire et syntaxe). Pour examiner le lien entre la perception de bas niveau, les traits de surfaces et les connaissances grammaticales plus abstraites, cette thèse s’articule en deux parties principales.

Dans la première partie, deux études longitudinales sont présentées. Chaque enfant était testé sur son seuil de discrimination auditive (avec un paradigme de traitement rapide) et sur une tâche de reconnaissance visuelle pour contrôler les compétences cognitives. Le seuil acoustique était évalué en utilisant des sons non-linguistiques (tons) dans un groupe d’enfants et linguistiques (syllabes) dans un autre groupe d’enfants à 9 mois pour déterminer si l’impact du traitement auditif est spécifique au langage. Enfin, le niveau de vocabulaire a été mesuré à 12, 14, 18 et 24 mois et un test cognitif (Mullen Scale) a été réalisé entre 18 et 20 mois comme supplémentaire contrôle cognitif. Les résultats montrent que les capacités mesurées sont prédictives du développement du vocabulaire chez les enfantes typiques. De plus, les enfants atypiques ont montré des capacités de traitement moins efficaces dans les modalités visuelle et acoustique.

Dans la deuxième partie, le rôle de la fréquence des mots dans l'amorçage des catégories lexicales des mots de fonction et de contenu et leur ordre relatif est explorée. Les deux catégories lexicales sont différentes dans leur fonction linguistique, leurs caractéristiques phonologiques et leur fréquence dans la parole. Ainsi, leur catégorisation basée sur la fréquence pourrait constituer un mécanisme initial robuste pour acquérir les constructions de base de la langue. Comme les mots de fonction constituent une classe fermée, alors que les mots de contenu constituent de classes ouvertes, nous avons examiné si les nourrissons français âgés de 8 mois étaient sensibles à la fréquence des mots pour catégoriser les mots de fonction et pour les traiter comme des éléments non-remplaçables dans des classes fermées, et les mots de contenu comme des éléments librement remplaçables dans des classes ouvertes. Les cinq expériences de grammaire artificielle menées ont confirmé cette hypothèse. De plus, les nourrissons associent l'ordre relatif de ces catégories avec l'ordre des mots de base de leur langue maternelle, le français. Les participants atypiques ont montré des capacités de discrimination, de codage et de mémoire inférieures à leurs pairs typiques.

En conclusion, ce travail a permis une meilleure compréhension des capacités de perception contribuant au développement du langage. De plus, cette thèse a identifié de potentiels marqueurs comportementaux pouvant servir à l'identification précoce des apprenants atypiques.

Mots clef: modèles de *bootstrapping*, traits de surface, capacités de traitements, variabilité individuelle, populations typiques et atypiques

Abstract

The purpose of this work is to explore how infants perceive surface features of language at different levels of processing. Specifically, in both typical and atypical populations, we examined 1) the processing of low-level auditory cues and its relationship with later language outcomes and 2) the sensitivity to word frequency to create abstract linguistic representations of lexical categories and their relative word order.

Accordingly to bootstrapping models learners are able to extract abstract, structural and hence directly unobservable properties of the target language from perceptually available surface cues in the input that correlate with the underlying structure. Indeed, infants are sensitive to certain acoustic and phonological properties of the speech input, which in turn correlate with specific grammatical/syntactic structures. In order to map the link between the perception of these low-level, surface cues and more abstract grammatical knowledge, this work is organized in two main parts.

In the first part, two longitudinal studies are reported. Each infant received an auditory discrimination threshold task (using the rapid auditory processing paradigm) followed by a habituation/visual novelty detection task used as a control for general cognitive skills. The auditory discrimination threshold was evaluated using non-linguistic (tones) sounds in one cohort of infants and linguistic sounds (syllables) in another cohort of infants at 9 months in order to investigate the language-specificity of the process within the auditory modality. Subsequently, infants' vocabulary was assessed at 12-14-18 and 24 months and a cognitive test (Mullen scale) was performed at 18-20 months as another control measure for early processing competence. Results show that early processing abilities are predictive of later vocabulary size in typical infants. Importantly, atypical participants exhibited slower and less efficient processing abilities in both visual and acoustic modality.

In the second part, the role of word frequency in bootstrapping the basic lexical categories of function and content words and their relative order is explored.

The two lexical categories differ in their linguistic functions, phonological makeup and frequency of occurrence. Thus, their frequency-based discrimination could constitute a powerful initial mechanism for infants to acquire the basic building blocks of language. As functors constitute closed classes, while content words come in open classes, we examined whether 8 month-old French monolinguals relied on word frequency to categorize and track functors as non-replaceable items in a closed class, and content words as freely replaceable items in open classes. In five artificial grammar-learning experiments we have found that infants treat frequent words as belonging to closed classes, and infrequent words as belonging to open classes and they map the relative order of these categories onto the basic word order of their native language, French, a functor-initial language. Importantly, atypical participants showed lower ability of discrimination, encoding and memory when compared to typically developing peers.

Overall this work contributes to a better understanding of the perceptual abilities that directly contribute to language development. Moreover, it proposes possible behavioural markers that can be potentially useful in the early identification of atypical learners.

Key words: bootstrapping mechanisms, surface cues, processing abilities, individual variability, typical and atypical populations

The present thesis has led to the following publications:

Marino, C., Gervain, J. (accepted). The impact of generative linguistics on psychology: Language acquisition, a paradigm example. Accepted with minor revisions in Acta Lingusitica Academica, a Journal of General Linguistic. The projected publication year is 2019.

Marino Caterina, Bernard Carline & Gervain Judit. (under review).Word frequency is a cue to lexical category at 8 months.

Marino Caterina & Gervain Judit. (under review). The novelty effect as a predictor of language outcome.

Acknowledgements

It seems impossible until it's done, but it wouldn't be possible without the contribution of many people.

Premièrement, je voudrais remercier toutes les familles et les enfants qui ont participé à mes études. Rien n'aurait été possible sans leur intérêt et leur motivation.

Thank you to Judit, not only a passionate thesis director but also a brilliant mind that has motivated me each day of these three years. Thank you for your optimism, for the scientific discussions, for the incredible knowledge that you are able to share so easily. Thank you for being open to any different interpretations, for having valorised all my ideas, even when difficult or incorrect and for having always been present but leaving room for the mistakes that have contributed to my improvement. You taught me that sometimes limits are only mental.

Thank you to the member of my Jury, LouAnn Gerken, Julie Franck, Anne Christophe and Jean-Rémy Hochmann. I am really honoured that you have accepted to read my thesis and be present the day of my defence.

Merci à Thierry, même si nous n'avons pas directement travaillé ensemble. Ta précision, ta détermination, tes connaissances et ton esprit critique contribuent toujours à rendre ce groupe non seulement une expérience scientifique forte, mais aussi une expérience de vie.

Merci aussi à tous les membres du LPP, particulièrement à l'équipe Parole pour avoir rendu cette ambiance de travail si agréable. En particulier à Viviane et Maxine, deux piliers qui font du BabyLab un endroit fonctionnel et accueillant. Toujours prêtes à m'aider, toujours prêtes à répondre à des questions pratiques et toujours prêtes à me réconforter dans les moments difficiles, avec un sourire, avec un bonbon à la violette ou simplement avec leur présence.

Gracias a mi Irenita, constante inspiración y fuerza. Gracias para no haber parado nunca de animarme, de escucharme, de reforzarme y de haber compartido conmigo todos los momentos de cotidiana locura intelectual (y no). Soy tu primera admiradora.

Thank you to the girls of the ex-608, now 627, particularly to Elena, Maria and Lisa for your support, for all the moment that we have shared, the breaks with gossip, the positive environment and the laugh.

Grazie a Mélanie, Irene, Ludo e Rosa, voi siete la mia Parigi. Merci pour ces années partagées de manière intense et quotidienne, merci pour la patience, pour m'avoir suivi dans toutes mes folies, pour avoir été présentes dans tous les moments, difficiles et importants. Merci pour m'avoir montré votre vraie âme.

À Mélanie, parce que cette ligne d'arrivée n'aurait pas été possible sans toi, sans ton support quotidien, ta patience, la bonne humeur de la maison, la musique horrible, le vin nature, l’ancien français, nos discours interminables, notre enrichissante diversité. Mais merci surtout parce que, en m'enseignant à ne pas réprimer les émotions qui me font sentir faible, tu m'as rendu beaucoup plus forte.
A Irene, perché dalla prima volta che ci siamo incontrate ho subito sentito il calore di casa. Grazie per il sostegno reciproco, per avermi insegnato a riflettere sulle mie potenzialità, per non aver mai smesso di rimproverarmi sulle mie mancanze e perché nella tua ordinaria follia io trovo la mia pace. Grazie soprattutto per avermi elencato tutte le malattie del mondo e per i tuoi vestiti che ho sempre adorato ma che non ho mai avuto il piacere di indossare.
Alla Ludo, scoperta tardiva ma non meno importante. Grazie per averci dato modo di recuperare così velocemente il tempo perduto, per esserti fatta conoscere in tutte le tue sfumature delicate, per avermi sostenuto da subito con dolcezza e con quel sorriso coinvolgente che rende colorato ogni contorno. E a Rosa, per le nostre risate compici, per essere l’unica a capire le mie battute, per la tua finezza nell’esprimere i concetti, la forza d’animo e la tua affascinante capacità di condividere le tue passioni.

Alla Bibi, la Carola e la Chiara (vi ho messo in ordine alfabetico), certezze di vita, legami imprescindibili. Siete da sempre il mio posto sicuro, la mia più grande forza e debolezza. Grazie per aver compreso l’imprevedibilità della mia assenza e il costante valore della vostra presenza. Con voi non ho bisogno di molte parole, voi siete la mia isola felice.
Grazie alla Carol che, seppure lontana, rende incredibilmente possibile condividere ogni momento significativo, alla Lu per le sue parole confortanti e le sue attenzioni sempre delicate e a Luca che da oltreoceano resta l’unico uomo capace di sopportare tutte le mie stranezze.
Grazie alle mie amiche di sempre, Gaia, Paola, Claudia, Emiliana, Michela, Chiara, Fabiola, Federica e Giorgia, perché il tempo e le diversità di vita non hanno scalfito la nostra bellezza. In particolare però voglio ringraziare Stefania, Francesca, Daniela, Irene e Marta per il sostegno, la dolcezza, la razionalità, la pazienza e il coraggio regalatomi in quest’ultimo periodo (non solo ma particolarmente).

Il più grande grazie è però per voi, il Pierlu, la Cri, Etienne e la Zazi. Una famiglia che auguro a chiunque. In questi ultimi anni ho imparato a vedervi non solo come membri di una famiglia ma come persone, osservando le vostre singole complessità, debolezze e punti di forza. E grazie a questo ho scoperto che dal nostro legame “biologico” nascono l’affetto e la condivisione quotidiana, ma è solo dal nostro legame come persone che abbiamo imparato a condividere un amore, una forza e un rispetto così autentici.

Et en fin, une remerciement à Paris, parce que pour moi tu ne représentes pas une ville mais un changement, une redécouverte. Je suis arrivée seule, perdue et incapable de communiquer mais aujourd'hui tu m'as donné une nouvelle famille, une nouvelle maison, une nouvelle langue, une nouvelle perspective et des satisfactions inimaginables. Dans tes murs, tu m’as rendue libre.

Take care of the sense, and the sounds will take care of themselves.

Lewis Carroll

Chapter 1. General Introduction

“Language is the expression of ideas by means of speech-sounds combined into words. Words are combined into sentences, this combination answering to that of ideas into thoughts.”
(Henry Sweet)

Acquiring language is one of the characteristics that makes human beings unique. This ability is established in infancy and it is observable long before spoken language emerges. Infants have a remarkable ability to perceive speech sounds. This ability rapidly evolves as infants gradually acquire knowledge of their native language by detecting and using different surface cues present in the speech input.

Decades of research have advanced several hypotheses to explain the general mechanisms guiding young learners to acquire language correctly and without explicit teaching. In this framework, *bootstrapping theories* proposed that learners are able to extract abstract, structural, and hence directly unobservable properties of the target language, from perceptually available cues present in the input that correlate with the underlying structure (Morgan & Demuth, 1996).

The present work explores perceptual learning processes relying on the surface cues of speech and investigates how they help infants bootstrap into language. As bootstrapping mechanisms are heuristic, infants may not exploit them in the same way or to the same extent. However, relatively little is known about the relation between an infant’s early efficiency to perceive bootstrapping cues and his/her later ability to acquire words, morphology, syntax or other aspects of language. A more precise understanding of the relation between early perceptual abilities and later language outcomes is necessary, as exploring this relationship might contribute to our understanding of the extent to which early processing measures are language-specific.

Language development is remarkable in that infants across different languages uniformly go through the same sequence of large developmental milestones. However, the rate and fine details of linguistic development may be substantially variable across learners. A growing body of research has started to emphasize the importance of investigating this individual variability across learners. The investigation of individual variation might help to describe and predict individual trajectories more accurately. From a more clinical point of view, having more precise measurements of individual variability in typically developing infants might help to detect early differences in atypical populations. Early perceptual difficulties, for instance, may be good predictors of delayed linguistic development often present in atypical learners.

In light of above, the current thesis has different objectives. Most generally, it tests bootstrapping mechanisms at different levels of linguistic description. As speech is an acoustic signal, the thesis evaluates whether the speed and efficiency of auditory processing abilities play a role in how accurately speech is perceived (Chapter 2). This in turn might be potentially related to later linguistic knowledge/outcomes. Specifically, it tests the hypothesis that infants who are more efficient auditory processors are also faster at learning language. Infants' ability to discriminate acoustic information starts already in the womb. This early perception contributes to a cascade of events involved in the anatomical and functional maturation of the auditory system. Hence, individual differences in early auditory discrimination might directly influence the way sounds are perceived, which, in turn, might contribute differentially to later language outcomes.

In addition to acoustic cues, the speech signal also contains statistical regularities. These often correlate with underlying grammatical structures. The thesis investigates how infants use a basic statistical property, word frequency to establish the basic lexical categories of function and content words and their relative order, i.e. the basic word order of the native language (Chapter 3).

The thesis will examine these auditory and frequency-based learning mechanisms at the group as well as the individual level in typically developing infants, relating them to language outcome, specifically vocabulary size at 12, 14, 18 and 24, months. Furthermore, it will test the hypothesis that atypical populations show early difficulties in the use of these mechanisms. For this reason, in both the psychoacoustic and the frequency-based bootstrapping studies, a group of infants at risk for language impairment was also tested in addition to typical infants. Early difficulties on these tasks might help to explain the delayed linguistic behaviour often present in atypical learners. Moreover, these early difficulties might be used as early behavioural markers of language delay, allowing for a more timely intervention.

1.1 General theories of language acquisition

More than 50 years of language acquisition research has attempted to explain the mechanisms that allows infants to extract sounds, syllables and words from the speech and to acquire the native syntax, morphology, etc. Theoretical debates about language acquisition have been guided by two opposing views: empiricism and nativism.

In the empiricist camp, the behaviouralist view (Skinner, 1957) proposed a domain-general learning of language, where infants imitate what they have in the environment (actions, words, etc.). Skinner suggested that reinforcement and punishment shape infants' "verbal behaviour", leading to language acquisition. In this framework language learning is viewed as a stimulus-response phenomenon.

Constructivist/usage-based theories of acquisition (e.g. Piaget, 1952; Vygotsky, 1962; Tomasello, 2000) assume that language development is influenced by interactions between the infant (or child) and his/her social and physical environment. Cognitive development is assumed to lead to the growth of language, viewed as a symbolic system that helps understand the external world. Within this framework, infants are assumed to construct abstract linguistic representations gradually, being sensitive to statistical (co-)occurrences present in the language input. For instance, from the sentences: "do you *have*?" "do you *want*?" "do you *need*?", infants might built a semi-formal representation of the linguistic construction "do you *X?"*. In this sense, learning is linked to specific lexical items and the general representation of structure emerges gradually through increasingly abstract representations. The general hypothesis guiding usage-based/constructivist theories is that "meaning is use and structures emerge from use" (Tomasello, 2003). Infants are equipped with two major cognitive structures: *intention reading* and *pattern finding*. According to the to the social-pragmatic approach, "intention reading" is meant as the ability to discover linguistic intentions that are related to the cultural or social context (Bruner, 1983; Nelson, 1996; Tomasello 1992; 2000). "Pattern finding", more often used by the usage-based approach, is considered as the ability to discover schemas and rules presented in the linguistic input (Goldberg 1995, 2006; Tomasello 2000, 2003).

In contrast with these theoretical positions, Chomsky (1959, 1965) advanced the idea that language acquisition has an innate basis (nativist theories). In his review of Skinner's *Verbal Behavior* (1959) he argued about the logical impossibility to derive a rule system from the finite set of linguistic exemplars that learners might encounter. Drawing a parallelism with the mathematical phenomenon of the "induction problem", suggesting that for any finite dataset there is an infinite number of underlying rules that could generate it, Chomsky argued that is impossible to know the specific rule systems that has to be selected to acquire native language grammar. In other words, there is no guarantee that two infants that are exposed to the same grammatical system would end up learning the same language.

This "poverty of the stimulus" problem is further exacerbated by the fact that the linguistic input that young learners receive often contains errors, omissions, speaker variation, and it is often perceived in a noisy environment. Moreover, infants not only never receive explicit teaching about the correct rules of their grammar, but they also rarely encounter negative evidence, i.e. information that a specific sentence is ungrammatical or it is not part of the native grammar (Marcus, 1993). It is, therefore, impossible for them to reject hypotheses about what rules are not present in their native language.

Therefore, since language cannot be learned from external input alone, Chomsky proposed that we have an innate language faculty grounded in our genetic endowment. Specifically, humans are equipped with a biologically determined language faculty, the "Universal Grammar" (UG), even prior to linguistic experience (Chomsky 1986). UG contains the logical space that defines all possible grammatical regularities in human language. The "language acquisition device" allows infants to learn their target language starting from a universal initial state of UG on the basis of limited experience.

UG contains sets of *principles* that hold universally (thus shared across languages). Language differences then depend on the values of the *parameters*, or binary switches that account for cross-linguistic variation (Principles and Parameters model, Chomsky, 1981). A particular grammar is characterized by a specific setting of the parameters, and their combinations account for *language-specific syntactic systems* (Chomsky, 1980, 1981; Lightfoot 1991; Rizzi, 1986).

Related to the idea of innateness, Lenneberg (1967) emphasized that language is profoundly constrained by biology. He argued that language is learnable only during a fixed time during development. During this period, the interaction of environmental and maturational mechanisms leads to native language competence. More specifically, this critical (or sensitive) period is considered the maturational time window during which the linguistic system is open to modification on the basis of external input (Bateson, 1979; Scott, 1962; Scott & Martson, 1950; Hensch, 2005; Bateson, 1979; Michel & Moore, 1995). Different aspects of language have different critical periods (Werker and Hensch 2015). Phonology, grammar and the lexicon develop at different periods during the life span (Kuhl, 2010). The temporal window for syntax closes around 7 years (Johnson & Newport, 1989), while for native phonemic attunement around 3-4 years (Werker & Hensch, 2015). While the critical periods for language close before puberty, recent research suggests that neural plasticity may be reopened during an individual's whole life span under the influence of relevant experience or certain drugs (Bavelier and Davidson, 2013).

1.2 Models of bootstrapping

Following the Principles and Parameters model (P&P; N. Chomsky 1981), Pinker (1984) introduced the term "bootstrapping" into language acquisition research. He argued that grammatical rules and categories are abstract, thus they are not directly available in the speech input. How does an infant know how to link these abstract mental contents to the acoustic input he/she receives? Pinker (1984) proposed to solve this *linking problem* by assuming an innate correspondence between semantics and syntax. In the *semantic bootstrapping hypothesis*, he proposed that basic semantic notions present in our everyday life (e.g. "actions" or "concrete objects") could be directly linked to syntactic or lexical categories (e.g. " verbs" and "nouns"). He assumed infants to have innately specified information about nouns referring to objects, and verbs referring to actions. By observing the contingency between a specific word and its meaning, infants might be able to link semantic information to syntactic structure (*linking rules*).

Nevertheless, it has been argued that semantic bootstrapping is not always possible. This mechanism requires infants to process the structure of the whole sentence to establish the correspondence with semantics. The *syntactic bootstrapping hypothesis* (Gleitman, 1990) proposed the inverse mechanism: it is by observing the syntactic structure that infants may deduce knowledge of the meaning. Specifically, this procedure "deduces the word meanings from the semantically relevant syntactic structures associated with a verb in input utterances" (Gleitman, 1990 p. 30). For instance, a transitive verb is likely interpreted as an action performed by an agent on an object (e.g. *he sent a letter*), whereas an intransitive verb is more typically an action performed by an agent with no object involved (e.g. *she was laughing*). Using this regularity, infants may categorize a verb as transitive if it appears in the sentence with two nominal arguments (in the example: *he*: Subject, *a letter*: Object), but as intransitive if only one noun phrase is present. This is a plausible mechanism to learn verbs, because infants learn nouns earlier than verbs. Indeed, infants' first 50-100 words mainly consist of nouns. The syntactic bootstrapping hypothesis does not deny the existence of the semantic one, but it proposes that the two mechanisms act in combination rather mutually exclusively.

Infants are sensitive to the acoustic information carried by the speech stream from very early on. The central argument of the *prosodic/phonological bootstrapping hypothesis* (Gleitman & Wanner, 1982; Nespor et al., 2008; Morgan & Demuth, 1996) is that certain acoustic/phonological properties of speech cue structural properties of syntax. Variations in duration, intensity and pitch of the speech sounds convey the prosodic hierarchy, which accounts for the systematic, hierarchical structure of how smaller phonological units combine into larger ones, e.g. how words combine into phrases and phrases into utterances (Nespor & Vogel, 1986; Selkirk, 1984; Nespor et al., 2008). These, in turn, map onto syntactic units, although the mapping is not perfect. Therefore, sensitivity to acoustic variations carrying prosody might help infants to parse speech into smaller, syntactically relevant units (Morgan & Demuth, 1996; Christophe et al., 1997; Christophe et al., 2003).

The syntactically most relevant prosodic constituents are the Phonological Phrase (PP) and the Intonational Phrase (IP) (Nespor & Vogel, 1986, 2007). The Phonological Phrase is the "layer of prosody immediately beneath the intonational phrase that combines prosodic words and clitic groups into a single prosodic unit" (Nespor & Vogel 1986, p. 165-186). The PP corresponds to the content word head of a syntactic phrase and its associated function words on its non-recursive/non-branching side (e.g. [*the cars*] [*that I saw*]. The Intonational Phrase dominates the Phonological Phrase and corresponds to the intonational contours (Pierrehumbert & Hirschberg, 1990). The Intonational Phrase is characterized by an initial increase in pitch, which then gradually declines (pitch declination). In line with all the linguistic hierarchies, while the number of Phonological Phrases contained in an Intonational Phrase may vary, Phonological Phrases never occur across Intonational Phrase boundaries (Nespor & Vogel, 1986, 2007). Smaller units are always exclusively and exhaustively contained in larger ones.

One domain in which prosodic bootstrapping has been well established is the acquisition of word order. Languages of the world vary systematically in the relative order of their principal syntactic components such as the Verb (V) and its Object (O) and more generally functors and their corresponding content words (see examples below).

The phonological phrase in VO languages (e.g. French, English and Italian) is characterized by final prominence, marked by increased duration on the prominent item as compared to the non-prominent item (e.g. *to* ***Ro:me***). OV languages (e.g. Turkish, Japanese and Basque), by contrast, are characterized by initial prominence, marked by increased pitch or intensity on the prominent as compared to the non-prominent element (e.g. Japanese: ***^Tokyo*** *kara* Tokyo from 'from Tokyo'). Languages with different word orders thus use different acoustic cues to mark prominence.

Experimental evidence has indeed demonstrated the validity of the prosodic bootstrapping model. Infants between 6 and 12 weeks were tested with a modified version of the non-nutritive sucking procedure, evaluating habituation/dishabituation rate. They were presented with pairs of sentences in French and Turkish. The two languages differ in the order of syntactic components. French is a VO language (*la dame lit un livre* 'the woman reads a book'), whereas Turkish is an OV language (*kadin kitabi okudu* the woman the book read). The sentences were presented to infants in a delexicalized form where all vowels were replaces by schwas, and consonants by a given member of their respective category of manner of articulation, thereby suppressing segmental and lexical information, but preserving prosody. Infants were able to distinguish between French and Turkish sentences using the prosodic cues alone. Authors argued that access to the syntax might be guided by prosodic information carried by the speech signal (Christophe et al., 2003).

Furthermore, 8-month-old OV-VO bilinguals (English-Japanese) were able to use these acoustic cues to guide their choice of word order (Werker & Gervain, 2013). The authors tested this hypothesis using an artificial grammar task were an alternating sequence of frequent (A & B) and infrequent (X & Y) syllables, mimicking the statistical occurrence of functors and content words in natural language, was presented. By concatenating these four-syllable-long basic units (AXBY) repeatedly, a speech stream was created. One group of bilinguals was familiarized with an OV prosody in which infrequent words (X & Y) were higher in pitch with respect to frequent ones (A & B). A second group of bilinguals was exposed to a VO prosody in which frequent words (A & B) were shorter than the infrequent ones (X & Y). In both conditions, bilinguals showed the expected preference for the word order that corresponded to the prosody. This suggests that bilingual infants are highly sensitive to prosody as a cue to word order. However, prosodic bootstrapping requires familiarity with the prosody in the target language. English monolinguals showed no word order preference after being familiarized with the OV prosody condition, as OV prosody is unfamiliar to English infants, exposed to a VO language.

Another well-established cue that can be used to bootstrap word order is word frequency. As mentioned above, languages conform to a basic word order type, characterized by the relative position of syntactic elements such as the Verb (V) and its Object (O) and function words (articles: *the*, adpositions: *on* and pronouns: *her*) with respect to content words (nouns: *giraffe*, verbs: *bring* and adjectives: *scary*). The *frequency-based bootstrapping hypothesis* (Gervain et al., 2008; Bernard & Gervain, 2012) relies on the language universal division of labor between functors and content words (Chomsky, 1995; Fukui 1986; Abney, 1987). The two lexical classes differ in their statistical occurrence. Individual functors are much more frequent than most individual content words (Cutler & Carter, 1987; Cutler, 1993; Kucera & Francis, 1976; Gervain et al., 2008). Therefore, in functor-initial, VO languages, syntactic units typically start with a frequent word (e.g. English: ***From*** *London*; Italian: ***Da*** *Londra*; French: ***De*** *Londres*), whereas in functor-final, OV languages, they end in a frequent word (e.g. Japanese: *Tokyo* ***kara*** (Tokyo from). Sensitivity to this relationship between word frequency and word order is detectable pre-lexically. 8-month-olds exposed to languages with opposite word orders, e.g. functor-initial Italian and functor-final Japanese, showed opposite preferences for word order in an artificial grammar task. Italian infants preferred sequences starting with a frequent word, while Japanese infants preferred sequences starting with an infrequent word, mirroring the word orders of these two languages (Gervain et al., 2008).

Table 1.1. *Example of the alignment of different cues to word order in English (VO, frequent-initial) and Turkish (OV, frequent-final).*

English:	***The***	***girl***	***is***	***eating***	***an***	***apple***
Prosodic information	Non-prominent	Prominent	Non-prominent	Prominent	Non-prominent	Prominent
Syntactic information	Functor	Content	Functor	Content	Functor	Content
Frequency information	Frequent	Infrequent	Frequent	Infrequent	Frequent	Infrequent
Word order	Subject		Verb		Object	
Turkish:	***Balık***		***limon***	***Ile***	***yenir***	
Prosodic information	Prominent		Prominent	Non-prominent	Prominent	
Syntactic information	Content *(fish)*		Content *(lemon)*	Functor *(with)*	Content *(is eaten)*	
Frequency information	Infrequent		Infrequent	Frequent	Infrequent	
Word order	Subject		Object		Verb	

At the phrasal level, prosodic and word frequency information is aligned (Table 1.1). In VO languages, prosodic prominence falls on the final constituent of the phrase, which is typically a content word, whereas its functors, which precede it, are non-prominent (e.g. French: *J'aime les pommes: I love apples*). In OV languages, prominence also falls on the content word, but in these languages content words tend to be phrase-initial (e.g. Japanese: *Taroo ga tegami o kaita*: Taroo.nom letter.acc wrote) (Nespor et al., 2008, but see also Gervain et al., 2008). Infants integrate word frequency and prosodic cues coherently. French monolingual infants (exposed to an artificial grammar similar to that used in Gervain & Werker, 2013) prefer the frequent-initial pattern (coherent with French word order) only when the prosodic and word frequency cues are aligned at the level of lexical items, i.e. the frequent and prosodically non prominent word, i.e. the "functor", appears in an initial position). Nevertheless, they show no word order preference when the two cues are misaligned, i.e. the frequent words are prosodically prominent). This result confirms, first, that infants expect prosodic cues and word frequency to be aligned at the lexical level. Moreover, it demonstrates that they process the two cues simultaneously (Bernard & Gervain, 2012).

Bootstrapping mechanisms rely on the assumption that abstract, structural properties of language are extracted from perceptually available surface cues. The correlation between abstract features and surface cues is typically probabilistic, not perfect. Some functors, for instance, may be less frequent in the input infants receive than some content words (e.g. *beyond* vs. *cup*, *eat*, *bottle*). Hence, these mechanisms are heuristic.

1.3 Individual variation, different developmental trajectories

Bootstrapping approaches exploit the multiple, partially redundant surface correlates of grammatical structure. These cues are different in nature (acoustic, statistical etc.) as well as in how reliably they cue the lexicon or grammar. It is, therefore, possible that different infants rely on these cues to different degrees. The efficiency with which they process the cues may also differ across infants, with some of the very inefficient learners possibly even suffering delay in the acquisition process. The thesis, therefore, investigates individual variation across infants in these tasks both within the typical population as well as comparing typical and atypical infants.

Most developmental studies either investigate early speech perception or later language outcomes, but do not link the two directly. Only recently did longitudinal studies start to link early assessments of speech perception abilities and later language outcomes. Several studies (e.g Tsao, Liu & Kuhl, 2004; Cristia & Seidl, 2011; Newmann et al. 2006, Fernald et al., 2006) have found supporting evidence about a direct link between infants' abilities to perceive acoustic cues and later expressive and receptive vocabulary. Acoustic abilities might in fact diverge at the individual level, possibly leading to different rates of linguistic development. We will thus take a longitudinal approach testing perceptual abilities early, and following infants up until age 24 months. To better understand the cascading and interrelated effects of experience and maturation during development, and how individual differences in early abilities may induce diverging trajectories during this process, we will briefly review the development of the auditory system and early auditory abilities.

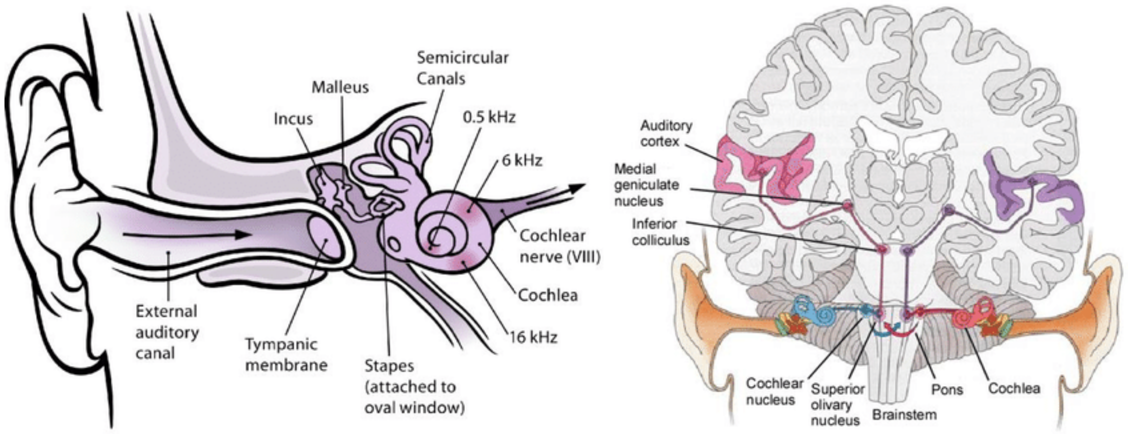
1.4 The basic principles of hearing and the development of the auditory system

The perception of the acoustic variations in pitch, duration and intensity that carry prosody is in place since the earliest stage of the development, as foetuses experience prosodic cues already at birth.

The role of the auditory system is to process the acoustic signal and to segregate sounds from one another in order to build the "auditory world" (Litovsky, 2015). The auditory system develops early during intrauterine life and it is divided into two main sub-systems: the peripheral auditory system, formed by the inner, middle and outer ear and the central auditory system, which goes from the cochlear nucleus until the primary auditory cortex.

The outer ear collects sound. From outside, sounds enter into a tubular structure called the auditory canal that amplifies them. From here sounds reach the tympanic membrane (eardrum) in the middle ear. Within the middle ear, a chain of three bones (incus, malleus, stapes) connected to the tympanic membrane and the cochlea further amplify sound, transmitting the pressure of the vibrations to the oval window. Increasing pressure is necessary since the inner ear is not filled with air but with liquid. When the sound pressure is transmitted into the fluid it stimulates the basilar membrane in an area that is specific for the frequency of the vibration. Consequently, lower frequencies cause movement in the upper (apical) part of the cochlea, whereas higher frequencies are encoded in the lower (basal) part (tonotopic mapping). In the cochlea, mechanical information is converted into a neural signal. More specifically, the vibrations are converted into electric signals by two types of mechanosensory cells (inner and outer hair cells). The auditory signal is then transmitted by the cranial nerve VIII. This nerve is composed of the auditory nerve that transmits sound-induced neural information and the vestibular nerve that serves for balance. The neural signal is then sent to different subcortical and cortical brain structures in order to be processed. By crossing the midbrain it reaches the thalamus, then the primary auditory cortex in the temporal lobe. The primary auditory cortex processes acoustic cues such as pitch, duration, and intensity (Graven & Browne, 2008, but see also Kandel, 5th edition, 2012).

Figure 1.1. The peripheral (adapted from Chittka & Brockmann, 2005) and the central auditory system (adapted from Rosenzweig et al., 2005), from Chen, 1987, thesis dissertation.



During ontogeny, the development of the cochlea and the middle ear starts at 16 weeks of gestational age. Between 25 and 29 weeks gestational age, the auditory system becomes functional as the first neural connections along the auditory pathway are established (e.g. Pujol and Lavigne-Rebillard, 1992). Thus, the anatomical and the physiological development of this system are guided by programmed changes, which in turn are influenced by the auditory stimulation experienced already in the womb (Litovsky, 2015).

Within the womb hearing is possible but sounds are low-pass filtered by maternal tissues. The acoustic signal is attenuated in frequencies higher than 600-1000Hz, thus much of the information characterizing the segmental level are cut out. Nevertheless, supra-segmental, i.e. prosodic information is preserved (e.g. Griffiths et al., 1994). As a result, prosody is perceived prenatally, suggesting that the intrauterine experience favours processing of prosody over the other attributes of speech (Abboub et al. 2016).

1.4.1 Processing the acoustic signal: early sensitivity to pitch, intensity and duration

Foetuses process the speech signal and respond to sounds (e.g. Hepper, 1991). Understanding how infants process acoustic information before and after birth might help to investigate how processing efficiency develops during the first years of life and whether and it is directly linked to language learning/outcomes.

Speech is a fast changing signal. It therefore requires efficient processing, otherwise incoming information might be lost. It is therefore possible that infants, whose auditory processing is less efficient than necessary, suffer delays in processing the linguistic input and as a result might be delayed or atypical in language acquisition.

The frequency of sounds is one of their most crucial characteristics, particularly relevant for speech. Frequency conveys a large number of different types of linguistic information as the identity of individual segments, especially of vowels, lexical pitch accent or tone, prosody etc. The human hear is able to perceive frequency variations that range between 20 and 20000 Hz. Importantly, the acoustic information carried by the human voice in the speech signal is smaller. The fundamental frequency of speech, called F_0, is between 300-350 in children, between 140-240 Hz in female speakers and between 100 - 150 Hz in male speakers. The bands that carry relevant acoustic energy for the identification of vowels, called formants, range between 250 and 4500 Hz. The frequency range relevant for consonant identification ranges between 100 and 8000 Hz.

The foetus is sensitive to low frequency sounds from 19 weeks of gestational age. Hepper and Shahidullah (1994) investigated the response to pure tones (from 100 to 3000Hz) between 19 and 35 weeks of gestation. They placed loudspeakers on the maternal abdomen and monitored the foetus's movements with ultrasound. Foetuses responded to low frequency tones at 100 and 250Hz already at 19 weeks. Nevertheless, it was only at 33 and 39 weeks, respectively, that foetuses started to respond to high frequency tones at 1000 and 3000Hz. The authors argued that the different responses to low and high frequencies reflected different maturational stages of the auditory system, as well as the filtering effect of the womb.

Similarly, a number of studies have revealed an enhanced sensitivity to low frequency ranges in newborns (e.g. Fox & Stapells, 1993; Lenard et al., 1969). Newborns' frequency discrimination threshold for pure tones was evaluated using frequencies varying from 100 to 1000Hz. In general, newborns showed greater sensitivity for low-frequency ranges, especially between 200 and 500Hz (Hutt et al., 1968; Eisemberg, 1970). Furthermore, a greater sensitivity to 1000Hz than to 4000Hz was observed in newborns between two and five weeks of age (Werner & Gillenwater, 1990). Hence, the ability to discriminate changes in frequency seems to be present already at birth (see also Nazzi, Floccia & Bertoncini, 1998). Nevertheless, it improves with development (e.g. Werker & Tees, 1983; Werker & Lalonde, 1988).

Sensitivity to duration has also been found at early stages of development. For instance, 2-month-old infants are able to discriminate non-speech sounds that vary in duration by about

20ms (Jusczyk, Myers et al., 1983). Between 6 and 8 months they can discriminate silent intervals of about 80ms between two tone sequences (Thorpe et al., 1988), silent gaps of 30-40ms between two identical sounds (Smith et al., 2006) and silent intervals between pairs of tones as short as 11ms (Trehub et al. 1995).

Early sensitivity to intensity was also documented. Seven- and 9-month-olds are able to detect intensity increments between 3-12dB (Sinnot & Aslin, 1985).

In general, sensitivity to the basic acoustic dimensions of sounds is present very early, even before birth. Importantly, intensity, pitch and duration vary systematically in the speech stream and contribute to the differentiation of prominent and non-prominent elements both at the word and the phrasal level. Therefore, early perception of these acoustic cues it is not only important in order to process psychoacoustic features of language but it might be directly related to language outcomes.

1.5 Is early perception linked to later language outcomes?

1.5.1 Typical populations

In the last years, a growing number of studies started to investigate the link between early perceptual abilities and later language acquisition. With a longitudinal design, Friederici and colleagues (2009) demonstrated that early discrimination of word stress patterns was correlated with language outcomes. Infants at 4 and 5 months were tested on their ability to discriminate native versus non-native word stress. ERP data were collected by measuring the MMR (mismatch response) with an oddball paradigm at 4-5 months. Native and non-native word stress patterns were presented either as standard and as deviant stimuli in two different conditions. The same infants were assessed in their language abilities at 2.5 years. Children that were later found to be high producers showed an early and late negative response when the native stress pattern was presented as a deviant stimulus and a positive response when the non-native pattern was presented as a deviant. In contrast, children with low expressive abilities at 2.5 years showed the same positivity when the non-native pattern was deviant but no negativity for the native pattern in the deviant condition. Moreover, the positivity elicited by the non-native pattern was longer in low than in high producers.

In addition, Cristia & Seidl (2011) showed that toddlers' vocabulary at 24 months was correlated with prosodic sensitivity at 6 months. Infants were tested with a HPP procedure where sequences of speech that were uttered as a well- or ill-formed prosodic unit were presented. Infants who preferred well-formed intonational units showed a larger vocabulary score (both receptive and productive) at 24 months with respect to those that did not show the same preference pattern.

Furthermore, investigating word segmentation abilities, Tsao, Liu & Kuhl (2004) examined discrimination of native and non-native vowels using a conditioned head turn procedure and its relation with later language outcomes (at 13, 16 and 24 months). They found that the infants who reached the conditioning criterion more rapidly showed a larger vocabulary score in comprehension at 13 months and in production at 24 months. In a similar study, the authors demonstrated that a similar effect was present when native and non-native consonant contrasts were measured. The discriminative performance at 7 months was correlated with expressive vocabulary at 18 and 24 months (Kuhl, Pruitt et al., 2005). This specific effect was confirmed by electrophysiological measurements. By comparing ERP responses for native and non-native speech contrasts, infants who at 11 months showed a discriminative pattern for the two stimuli had larger vocabulary at 30 months (Kuhl, Rivera-Gaxiola et al., 2005).

The relation between speech perception and vocabulary was also examined by classifying high- and low-vocabulary children at 2 years and examining correlations with prior performance in speech perception tasks (Newmann et al., 2006). Speech segmentation abilities measured before 12 months were correlated with expressive vocabulary at 24 months. Not only, the positive correlation between word segmentation skills, measured at 7.5 months and later vocabulary size (24 months) was replicated in English-learning participants (Singh et al., 2012).

In order to explore the performance of toddlers, English-learning participants at 15, 18, 21 and 25 months were tested in a looking-while-listening paradigm. In this paradigm two images (or more) were simultaneously presented while only one was named. Eye movements were measured at different ages as an index of speech comprehension and it was analysed in relation to vocabulary growth over the 2nd year of life. A growth curve analysis showed that children who were faster and more accurate during the task also had faster growth in the later expressive vocabulary. This result was particularly strong for the 25-months group. The online eye movements, measured by speed and accuracy towards the correct picture during the task, were strongly related to the lexical and grammatical development across ages (Fernald et al., 2006).

1.5.2 Atypical populations

The evidence provided above focused on typically developing participants, with no hearing or visual impairment and no familial history of language impairments. Although while a variety of studies have found/proposed early markers of language delay/impairments, a more clinical approach is particularly challenging. First because, during the first year of life, speech perception phenomena are still under development. Thus, assessing reliable early predictors might not be an easy task. Moreover, populations at risk are rarely homogeneous as they commonly present comorbidities. For instance, developmental dyslexia (DD) presents comorbidity with attention deficit hyperactivity disorder (ADHD) and speech sound disorder (SSD) (Pennington, 2006). Hence, isolating a single predictive factor is challenging.

Notwithstanding, many studies reported significant differences in speech perception abilities between typical and atypical populations providing a better understanding of these multiple factors. Below, studies showing delayed perceptual abilities in populations at risk for language impairments are reported.

1.5.2.1 Atypical populations: evidence from infancy

Friederici and colleagues (2004) explored auditory processing abilities in 2-month-olds at risk for specific language impairment (SLI). In a passive oddball paradigm, mismatch responses were measured by presenting CV syllables with changes in vowel duration. Infants at risk for SLI exhibited a delayed latency in the mismatch response for the deviant stimuli as compared to infants without risk.

Similar results were obtained when infants at risk for developmental dyslexia (DD) were tested at 6 months on consonant durational changes. They presented a delayed brain response to durational change as well as a poorer detection of stimulus changes (Leppäen et al, 2002). Consistently, when the same population was tested behaviourally, increased time to categorize speech sounds was measured at 6 months than in typical peers. Interestingly, the same difference was reported when testing their dyslexic parents in a similar, but age-adapted task (Richardson et al., 2003).

Overall, it seems that processing durational information and detect durational changes is significantly impaired in both the SLI and DD at risk populations. As duration is one of the relevant acoustic prosodic features in many languages, impaired perception of this cue might affect the perception of prosodic patterns and, in turn, directly interfere with the language acquisition process. Studies investigating prosodic abilities as early predictors of SLI seem in line with this perspective.

In a recent study, Höhle and colleagues (2014) followed infants longitudinally, testing them at 4 months and at 5 years. In the group, eight infants at risk of SLI were present. At 4 months, the processing of prosodic information was tested with an HPP procedure. German-learning infants were compared in their discrimination ability for the native trochaic (high-low) pattern versus the non-native iambic one (short-long). In the iambic stimuli, the second syllable was longer, whereas in the trochees, the first syllable was increased in intensity and pitch. The performance of the two populations was analysed separately. While typically developing infants showed the predicted preference for the trochaic pattern, no preference was found in the group of infants at-risk. Moreover, at 5 years, children's language performance was evaluated using a standardized German language Test (SETK3–5, Grimm et al., 2001) together with the Snijders-Oomen Nonverbal Intelligence Test (SON-R 2,5–7; Tellegen et al., 1998) to measure cognitive outcomes. Correlational patterns revealed a strong link between early rhythmic discrimination and later outcomes. Specifically, the absence of a family risk and a shorter looking time to the iambic stimuli were associated with higher vocabulary score (sentence comprehension and morphological abilities). Moreover, infants considered at risk were delayed in their linguistic, but not in their cognitive abilities.

Freiderici, Weber et al. (2005) performed a retrospective longitudinal study. Infants were first tested at 5 months in their prosodic abilities, and they were later split in two groups based on their language performance at 12 and 24 months. To test prosodic abilities two CVCV words with different stress patterns were presented using a passive oddball paradigm. Two conditions were tested. In the first condition the iamb version of the CVCV word was presented as the standard stimulus (occurring the 5/6 of the time) whereas the trochee version was presented as the deviant (1/6 of the time). In the second condition the two were inverted. Results were analysed separately for the two groups. Infants with lower productive abilities at 12 and 24 months had a family risk for SLI. In this group a negative deflection, at 300ms, was found for the trochaic deviant condition, whereas no deflection was present for the iambic one. The control group exhibited a similar pattern where negative deflection at 240ms was found for the trochaic deviant condition but no response was elicited by the iambic deviant one. Nevertheless, the comparison between the two groups revealed an increased negative deflection for the control group. Moreover, the amplitude of the negative peak was correlated, in both groups, with subsequent linguistic outcomes.

Importantly, as many acoustic processes in language occur fast, the ability to process the acoustic signal require efficient processing. For instance, in occlusive consonants, critical acoustic transitions (from occlusion to the release burst) often take place between 0 and 50 ms. Processing these rapid transitions is important for language comprehension (Aslin, 1989). The ability to process stimuli presented rapidly and sequentially is known as Rapid Auditory Processing (RAP). This ability involves the discrimination of two (or more) sounds presented one after the other, and can be operationally measured by determining the minimum interval between the two sounds that is required for successful discrimination. The rapid auditory processing paradigm (RAP) has been extensively used to predict subsequent language outcomes. Moreover, it has also been proposed as an early behavioural marker of language impairments. By testing 7.5-month-olds, it has been revealed that the mean acoustic threshold was higher (thus worse) in infants with a family risk for SLI. In addition, the individual acoustic threshold in both populations was correlated with subsequent linguistic outcomes (Benasich and Tallal, 1996; 2002).

Decades of research on language-impaired population have extensively reported impaired abilities on tasks requiring a rapid integration of sounds (see Tallal et al., 1993, Benasich and Tallal, 1996; 2002). In the "the auditory temporal deficit hypothesis" (Tallal, 1984; Tallar & Curtiss, 1990), the authors claimed that infants whose acoustic performance diverges from the norm, are more likely to be impaired or delayed in later language acquisition (see also Tallal, 2004). This hypothesis argues that difficulties in acoustic processing might impair the phonemic mapping (the perception of phoneme category membership) within the native language. Electrophysiological studies have also demonstrated that impairments in rapid auditory processing have signatures at the neural level (e.g. Richards & Goswami, 2015; Guttorm et al., 2001; 2003; 2005; Benasich et al., 2006; 2002).

Slower processing not only interfere with low-level discrimination but it might also result in difficulties in creating representations necessary for lexical and grammatical competences (Leonard et al., 2007). Nowadays, the auditory temporal deficit is used as a reliable marker for the early detection of language impairment. Overall, early psychoacoustic measures seem strongly reliable when individual variability wants to be assessed. For this reason, a more detailed description of theoretical and experimental evidence supporting the importance of rapid auditory processing abilities in language acquisition is discussed in chapter 2.

For instance, some evidence in the literature has revealed that auditory processing measures were correlated not only with later language outcomes but also with more general cognitive skills (Benasich and Tallal, 1996; 2002). Hence, the proposition about the involvement of multiple factors (non only linguistic) when language processes are evaluated was confirmed. Relevantly, the importance of controlling for general cognitive measures when linguistic abilities are evaluated is an important issue presented in chapter 2.

1.5.2.2 The atypical population tested in the current thesis: infants at risk for developmental dyslexia

In this thesis, we focus on infants at risk for language impairment as our atypical population. Several reasons guided this choice. First, language impairment impacts language performance from infancy but it is clearly observable only from the pre-school age. Hence, assessing early perceptual abilities might lead to a more specific understanding of the differences between typical and atypical development during the first year of life and might provide early markers of potential language delay useful for clinical purposes. Moreover, retrospective studies supported a genetic base for language impairments showing a high degree of family aggregation (e.g. Hurst et al., 1990). Hence, the risk to develop language impairment is higher in infants that have parents and/or siblings already affected.

Importantly, the sample of the atypical population tested in the present thesis was homogeneous, including infants at risk of developmental dyslexia. Several studies have already demonstrated that dyslexic children and adults have poorer rhythmic and prosodic perception.

Developmental dyslexia (DD) "is a childhood learning difficulty that is defined as a specific difficulty in reading and spelling that cannot be accounted for by low intelligence, poor educational opportunity or obvious sensory/neurological damage. The core cognitive difficulty in developmental dyslexia lies with phonology, as measured by the ability to reflect on the sound structure of words" (Snowling, 2000, from Goswami et al., 2014, p. 262).

The genetic and neurobiological components of DD have been well studied. Even if there isn't a full agreement in the scientific community, the influence of genetic transmission has been consistently reported. Dominant transmission occurs in a significant number of dyslexic families (Pennington et al., 1991) and the risk of infants born in dyslexic families to develop dyslexia is 2-80 times larger than the population average (Gilger et al., 1991).

Moreover, it also seems to be the case that there is a specific neurobiological component. Studies with post mortem brains have revealed abnormalities in the left hemisphere between the planum temporale and the perisylvian region. Specifically, the medial geniculate nuclei (MGN) of dyslexic, supporting the auditory processing system, were found to be smaller and having a smaller number of neurons when compared to non-dyslexic brains (Galaburda & Kemper, 1979; Galaburda et al., 1985; Humphreys et al., 1990). Hence, the neuroanatomical abnormalities are aligned with the behavioural manifestation of DD.

Poor rhythm perception in DD is well established by many studies with children and adults. In a systematic review, Hämäläinen and colleagues (2012) have reported that auditory processing deficits are often present in dyslexics. Reviewing studies using different experimental techniques (from behavioural to neuroimaging), the authors concluded that frequency and duration discrimination thresholds, amplitude modulation, intensity and gap perception are impaired in individuals with dyslexia. Accordingly, Goswami, Gerson & Astruc (2010) reported significant predictive relations between auditory perception, measured as the amplitude of the speech envelope tracking response, prosodic sensitivity and phonological awareness in DD children. Furthermore, impaired sensitivity to syllable stress was measured in children with DD at 9 years as compared to typical peers (Goswami et al., 2013). Children with reading delays exhibited poorer abilities in discriminating both stop consonants and tones with short interstimulus intervals (Reed, 1989).

1.6 General objectives and research questions

In the light of above, the present thesis has a **double** objective. **It tests bootstrapping as a possible mechanism allowing infants to go from surface/acoustic features of speech to lexical and grammatical aspects of language**. Moreover, **it investigates whether these bootstrapping mechanisms vary across infants and whether they can be used as early behavioural markers of risk for language delay in atypical populations**.

The thesis consists of two main experimental chapters. In the first chapter (chapter 2), psychoacoustic measures of low-level acoustic cues are reported in two longitudinal studies. Specifically, the ability to process linguistic (syllables) and non-linguistic (tones) sounds presented rapidly and at different frequencies was investigated as potential predictors of later language outcomes. Repeated measures of language outcomes were collected to explore the predictiveness of processing abilities measured during infancy. Importantly, measures of general cognitive and attentional abilities were also taken and discussed in relation to both the acoustic and the later linguistic measures.

In the second chapter (chapter 3), bootstrapping mechanisms supporting more abstract structural learning are investigated. Specifically, the frequency-based bootstrapping hypothesis was tested to study native word order acquisition and the related ability to classify words into lexical categories. Five artificial grammar learning tasks were run in order to study early sensitivity to word frequency. Moreover, the additional role of prosodic cues is discussed.

Relevantly, in both chapters a comparison between typical and atypical populations is presented. In particular, the behaviour of infants at-risk for developmental dyslexia was investigated. This choice was made first, because poor/impaired rhythmic perception has been observed in many studies investigating dyslexic adults and children. Hence, the literature supports a link between several aspects of auditory processing like sensitivity to pitch, duration and intensity and impaired phonological mapping. Secondly, few studies in early infancy have investigated the role of auditory processing abilities in infants at-risk for developmental dyslexia, as studies mostly focused on SLI. Finally, to our knowledge, no existing study has evaluated bootstrapping mechanisms in this specific population at risk.

Chapter 2:

Testing acoustic and visual processing as predictors of later language outcomes: longitudinal evidence

2.1 Introduction

During the first years of life, studies have revealed a strong connection between early auditory perception and later language outcomes (e.g. Benasich and Tallal, 2002; Kuhl, Pruitt et al., 2005; Newmann et al. 2006). Early mechanisms of perception and encoding are thus developed well before spoken language emerges. One well-established measure of early auditory perception is the ability to perceive and categorize stimuli that occur fast (within tens of millisecond). These acoustic abilities and their neural substrates are in place from early in development.

More specifically, a growing body of research has shown that processing rapidly and sequentially presented auditory stimuli is a necessary underlying skill supporting language development (e.g. Benasich & Leevers, 2002; Benasich & Tallal, 2002; Fitch et al., 2001; Farmer & Klein, 1995). This ability, also known as rapid auditory processing (RAP), is crucial in order to process and segment speech accurately, and consequently it is also necessary for language comprehension.

Nevertheless, to date, still relatively little is known about the individual differences in auditory discrimination and their influence on later language outcomes. By performing a longitudinal study (following Benasich and Tallal, 1996), here we examined whether auditory processing abilities, evaluated at an early stage, precede and predict subsequent language outcomes.

In the following sections a review of different theories and experimental evidence is provided in order to support the proposal that early processing skills are in place since the earliest stage of development and they are later correlated with linguistic knowledge. In addition, as processing abilities have been suggested to be implicated in general perceptual-cognitive development, a visual novelty detection task was performed in order to control for general cognitive factors.

2.1.1 Processing efficiency, temporal acuity and speech processing

Speech occurs fast. When we hear someone talk we need to recognize as fast as possible the rapid auditory information that signals which word is being produced in order to understand a sentence. The majority of the sensory processes necessary for language comprehension and production occur in tens of milliseconds.

The ability to perceive the frequency of speech sounds is crucial in this regard. Frequency conveys a large number of different types of linguistic information: the identity of individual segments, especially of vowels, lexical pitch accent or tone, prosody etc. The human hear is able to perceive variations in frequency between 20 and 20000 Hz. The acoustic information carried by the human voice in the speech signal is somewhat less large. The fundamental frequency of speech, called F_0, is between 100 - 150 Hz in male speakers, 140-240 Hz in female speakers and 300-350 in children. The different formants of vowels, the bands that carry relevant acoustic energy for the identification of vowels, ranges between 250 and 4500 Hz.

Different frequencies within these ranges often need to be discriminated very fast. For instance, in occlusive consonants, critical acoustic transitions (from occlusion to the release burst) often take place between 0 and 50 ms. Such dynamic aspects of the speech signal suggest that these rapid transitional cues have to be perceived accurately in order to correctly process speech (Aslin, 1989).

The speed and efficiency of auditory perception may play an important role in how accurately speech is perceived, which in turn may play an important role in how well, how fast or how accurately language is acquired. Infants' ability to discriminate rapid frequency transition is necessary to efficiently categorize and discriminate the formant transitions critical for phonetic distinctions. Many studies in the literature have shown that these abilities are in place from very early on.

From the age of 3 months pure tone discrimination with frequencies ranging between 250 and 8000Hz has been measured using an observer-based psychoacoustic procedure (OPP). This is one of the first studies measuring absolute thresholds over broad frequency ranges at early stage (Olsho et al., 1988).

Moreover, when the processing of tones was evaluated, infants showed the ability to discriminate between multitone patterns with different temporal grouping already at 5 months. Infants' habituation/dishabituation rate was measured using heart rate while infants listened to pure tone sinewaves that differ in their contrasting temporal arrangements (Chang & Trehub, 1977).

In measuring infants' absolute auditory threshold discrimination, gap detection tasks have proven particularly useful as they strictly depend on temporal acuity. Temporal acuity, "the ability to resolve two sounds separate in time" (Irwin et al., 1985, p. 614) is a crucial mechanism in auditory perception that is in place since early infancy. For instance, when two rhythmic tone sequences were presented, infants between 6 and 8 months discriminated silent intervals until 80 ms (Thorpe et al., 1988). Moreover, a gap-detection threshold of about 11 ms was measured in infants between 6.5 and 12 months when presented with pairs of 500-Hz tones (Trehub et al. 1995).

Even earlier, at 2 months of age, using the high-amplitude sucking (HAS) procedure, infants presented with non-speech signals that mirrored speech properties, were able to discriminate differences in the onset of two elements. Discrimination was measured by presenting two-tone sequences of 500Hz and 1500Hz. Stimuli were arranged along a temporal continuum where the 1500Hz-tone had a constant duration of 230ms while the duration of the 500Hz-tone varied from 300 until 160ms (steps of 10ms) (Jusczyk et al., 1980).

By longitudinally examining infants between 6.5 to 9 months old on discrimination of frequency transitions, Aslin (1989) reported that, even if not as well as adults, they were able to discriminate upward and downward frequency sweeps quite accurately. The author argued that this specific ability "is critical to discrimination and categorization of phonetic distinction in natural speech" (p. 588).

The ability of discriminating between phonetic units is present early. For instance, 2- and 3-month-olds already process spectrally distinct, but categorically identical sounds as perceptually equivalent. After being trained with an observer-based procedure, infants learned to positively respond when changes in vowel category (e.g. /a/ to /i/) occurred and to not respond when vowel category remained the same but variations in the pitch and spectral cues were applied (Marean, Werner & Kuhl, 1992).

Taken together, these studies demonstrate that acoustic abilities needed to process rapid changes of speech effectively are present early. Nevertheless, acoustic processing improves during the first years of life (e.g. Elliott & Katz, 1980; Fior, 1972; Roche et al., 1978). Tracking this developmental trajectory is relevant not only to explain how processing efficiency improves across ages, but it also allows a better understanding of how this specific ability is directly related to language outcomes.

2.1.2 The developmental improvement of acoustic processing: evidence from infancy to adulthood

In studying the development of acoustic processing many studies have compared adults and children's performance using different paradigms. The existing literature shows that processing abilities improve with age and linguistic expertise.

Elliot (1979) has compared the intelligibility of speech and noise related to age. Adding noise reduces the amplitude modulation of the speech waveform. In this study infants between 9 and 17 years were presented with the SPIN test (Speech Perception in Noise), showing that the youngest group performed significantly worst than the older, moreover the performance improved with age.

This evidence was later confirmed by Neuman and Hochberg (1983). Phoneme identification of nonsense syllables was presented under two conditions of reverberation (monaural and binaural) to children between 5 and 13 years and to adults. Reverberation is an acoustic phenomenon that blurs the intensity of temporal fluctuations of speech waveform (Houtgast & Steeneken, 1973). Authors reported that children have a selective difficulty in processing reverberant speech and reach adult-like performance only at 13 years of age. The age improvement was interpreted as reflecting the development of temporal acuity in speech processes.

Davids and McCroskey (1980) tested a large cohort of infants between 3 and 12 years old. They measured the ability to determine the minimum separation detectable between two tones, which they called "auditory fusion". Specifically, they asked children to distinguish paired and single tone pulses with intervals that varied from 0 to 40 ms. Results across ages confirmed the hypothesis that there is a progressive decrease in time (thus better performance) in the individual thresholds.

A similar age effect was founded by Irwin et al. (1995). Here children between 6 and 12 years old were compared to adults. Using a 2-alternative forced-choice paradigm, temporal acuity was measured by determining the minimally detectable silent interval in a continuous sound (that changed in frequency and intensity). Gap detection improved significantly with age, matching adult-like performance around 11 years. Moreover, this effect was stronger at lower intensities and low-frequency noise. The age improvement was attributed to the general development of auditory sensory processes.

A study by Wightman et al. (1989) yielded contrasting evidence. Authors found that 6-year-olds' performance in gap detection was adult-like, particularly for frequencies between 400 and 2000 Hz.

Considering that those differences might be related to the tasks themselves, it is not easy to track precisely the maturation of temporal acuity. For this particular reason, researchers started to test infant subjects in similar paradigms. For instance, 6-month-old infants were tested in their ability to respond to changes in duration of repeated noise burst. Infants' performance was worst than that of 5-year-olds', which was in turn worst than that of adults.

In addition, 3-6 and 12 month-olds were compared to adults. Participants were exposed to different conditions in which they had to detect silent intervals of variable duration in masked and unmasked sounds. Younger infants (3 and 6 months) exhibited the poorest threshold, followed by 12-month-olds and then adults in the unmasked condition. By contrast, in the masked condition, the performance of the 3 groups of infants was similar. Nevertheless, 12-month-olds showed a peculiar pattern in one of the masked conditions (500 Hz.). Here the performance was overall better and more adult-like than in the other 2 groups, even if huge individual variability was reported (Werner et al., 1992).

Moreover, the developmental improvement was confirmed at the level of discriminatory abilities for speech sounds. The authors detected an increased age performance testing groups of infants with a visually reinforced discrimination task presenting speech sound contrasts like /sa/ vs. /za/, /sa/ vs. /va/, etc. (Eilers et al., 1977).

Overall these findings suggest that temporal acuity improves with age and reaches an adult-like level around 13 years of age. Nevertheless, some results reported evidence of similar performances between adults and infants in masked noise (e.g. Werner et al., 1992) or for specific frequencies (Wightman et al., 1989). These might suggest that the general mechanism operates similarly across ages, and the observed improvement is due to linguistic experience.

Importantly, most of the data reported above is group level data. Nevertheless, infants' acoustic abilities might vary drastically at the individual level. This in turn raises the question of whether individual differences in early processing abilities are detectable and whether they are related to later language outcomes.

2.1.3 Individual variability: better processors, better learners?

Individual variability is frequently untracked. Nevertheless, a growing body of evidence starts to emphasize the importance of individual variation in the auditory temporal threshold and how this variation might be linked to later language development. Originally, studies about individual abilities were conducted to investigate language delays/impairments or even pathological conditions.

One of the first studies that tried to link the processing of auditory temporal information to language outcomes was conducted with dysphasic children. Perceptual, motor and sensory abilities were investigated and linked to receptive language abilities. Results demonstrated that the perceptual variables requiring rapid temporal processing were most highly correlated with the degree of receptive language deficit in this population (Tallal, Stark & Mellits, 1985a). Moreover, rapid auditory perception and production tasks alone were sufficient to classify children as language-impaired or normal (Tallal, Stark & Mellits, 1985b).

Decades of research on language-impaired populations have extensively reported impaired abilities “on tasks that require rapid integration of two or more sensory events which enter the nervous system in rapid succession” (see Tallal et al., 1993, for review, but also Benasich and Tallal, 1996, p. 351).

One of the hypotheses regarding the possibility to detect language delays early in life, “the auditory temporal deficit hypothesis” (Tallal, 1984; Tallar & Curtiss, 1990), claims that infants whose acoustic performance diverges from the norm, are more likely to be impaired or delayed in later language acquisition (see also Tallal, 2004). This hypothesis considers that difficulties in acoustic processing might impair phonemic mapping. Thus, infants that are delayed in this particular ability are more likely to be impaired or delayed in later acquisition.

In addition, authors have also proposed a general deficit of processing rapid temporal sequences both in auditory and visual modalities. This hypothesis claims that sequential processing can be considered a domain-general ability directly influencing language development (Tallal, Stark & Mellits, 1985a). Faster processing allows operations to be performed more rapidly and efficiently. Contrarily, slower processing would interfere in extracting sequential information form speech, resulting in difficulties with building up grammatical and lexical representations essential for language development (Leonard et al., 2007).

This theory was extensively applied to understand the etiology and the basic mechanisms of language impairments. In a series of longitudinal studies comparing infants at risk for language impairment (like specific language impairment; SLI) with controls, it has been shown that early differences in acoustic processing might be predictive of delayed or impaired language acquisition (e.g. Benasich & Tallal, 2002, see Leonard, 1998 for a review).

More recent studies have provided evidence that these differences are also reflected in how the brain processes rapid sounds. Goswami and colleagues have revealed differential electrophysiological response amplitude to the speech envelope when children with SLI and controls were compared. Impaired children seemed to exhibit a lower performance in rise time and duration perception, both of which are crucially related to the rhythmical properties of speech (Richard & Goswami, 2015). In addition, more specific differences in the ERP-components were observed in studies testing populations of infants at risk for dyslexia at birth (Guttorm et al., 2001; 2003; 2005) and in infancy (Benasich et al., 2006).

The auditory temporal deficit may therefore be a reliable early marker of language impairment. Benasich and colleagues (1996) have investigated for almost a decade the predictive role of RAP in later language outcomes. In one of the first studies, 7.5-month-olds performed a rapid auditory processing task. Two populations, at risk for SLI and controls, were compared in order to detect early differences. Infants were tested with a conditioned head turn procedure, in which they had to discriminate between two-tone pairs and learn to associate each one with a head-turn (right or left). The stimuli were tones with a fundamental frequency of 100Hz (tone 1) and 300Hz (tone 2). Two stimulus sequences were created. In sequence 1, tone 1 was followed by tone 1, in sequence 2, tone 1 was followed by tone 2. During the test phase interstimulus intervals (ISIs) between tones varied between 500 and 8 ms, allowing the assessent of temporal thresholds. Authors found that infants at-risk exhibited significantly higher auditory temporal threshold than controls. This finding was in line with the hypothesis that the auditory temporal deficit might be used as early marker of language impairment.

In 2002, the same authors provided evidence that rapid auditory processing is predictive of later language outcome. Measures of the receptive and expressive vocabulary as well as general cognitive performance were evaluated administering, respectively, the MacArthur Communicative Development Inventories (CDI) (Fenson et al., 1993) and the Preschool Language Scale-3 (PLS-3) (Zimmerman et al., 1992) together with the Bayley Scales of Infant Development (Bayley, 1969) between 6 and 24 months and the Stanford Binet Intelligence Scale (Thorndike et al., 1986) at 24 months. The rapid auditory processing threshold was found to be the best predictor of language outcomes at 2 years. Crucially, in all the measures, infants at-risk performed worst than controls, exhibiting a lower score with respect to controls at the same age. This confirmed that early deficits in RAP precede and predict language delays (Benasich & Tallal, 2002).

Importantly, auditory threshold measures are also particularly useful to investigate individual variability in normal populations. Trehub and Henderson (1996) provide evidence about the relation between early auditory measures (at 6 and 12 months) and later language abilities. Toddlers who were better processors at early stages obtained larger productive vocabulary scores when tested between 16 and 29 months. This is in line with the hypothesis that being a more efficient processor may directly influence the development of language. Thus, auditory processing skills are not only important to detect early populations at-risk, but they also underlie differences in language abilities in the typical population.

Tallal and colleagues suggested that information processing might not be modality specific, as the deficit was observed both in the auditory and the visual modality (e.g. Tallal, Stark & Mellits, 1985a, 1985b). This leaves open the question whether processing abilities, and consequently processing deficits, are domain-general or specific to the auditory/speech domain. More generally, the role of general cognitive abilities in language acquisition is still under debate.

2.1.4 Controlling for general cognitive factors

A controversial issue in the field of language acquisition is the extent to which general attentional or cognitive abilities play a role in individual differences in early language outcomes. A more accurate assessment of language development may perhaps be obtained if general attentional and cognitive abilities are also taken into account in addition to linguistic factors.

The contribution of general cognitive skills to later language outcomes is controversial in the literature. Some authors consider that general cognition mechanisms, such as attention or memory, are not sufficiently stable during early development and thus a poor measure of individual ability at any later stage (Kopp & McCall, 1982; Bayley, 1949; McCall; 1979). Some studies have also proposed that there is a huge variability in the magnitude of the cognitive effects, which makes it difficult to link them directly to later outcomes (Bayley, 1969).

Other researchers, by contrast, have proposed that looking time measures, like the increased response to a novel stimulus, are good predictors of later intelligence (Sternberg, 1985; Sternberg & Berg, 1985; Bornstein & Sigman, 1986). In fact, visual attention constitutes one of the major sources of infants' knowledge of the world. Hence, looking time measures have been largely used to discover the cognitive mechanisms supporting infants' development (e.g. Thompson et al. 1991; Colombo et al., 1989; Rose et al., 2009). Differences in looking times reflect distinct underlying cognitive processes, which in turn may be linked to individual differences (see Aslin, 2007 for a review). Responding to novelty involves two fundamental aspects. First, there is a "motivational" aspect by showing interest in or attending to something novel. Second, there is an "information-extraction/memory" aspect that involves the ability to identify features necessary to encode novel information and compare it to older information (Sternberg & Berg, 1985).

The visual novelty detection task has been found to predict later cognitive abilities as well as language outcome in a few existing studies. In one of the first studies, early novelty preference was found to be related to later memory skills (e.g. Colombo et al., 1989). Subsequently, Thompson and colleagues (1991) administered a test of visual novelty preference (FTII) to American infants at 5 and 7 months. The infants were followed longitudinally and the Bayley Scale of Infants Development (Bayley, 1969) was administered at 12 months, whereas the MacArthur-Bates Communicative Development Inventories (CDI, Fenson, 1993) were administered at 24 and 36 months together with the Stanford-Binet and the Colorado Specific Cognitive Abilities Test at 36 months. They found that early novelty preference was highly correlated to IQ at 24 and 36 months (not at 12 months). Furthermore, it also predicted language skills at 36 months. More recently, Rose and colleagues (2009) used a large battery of tasks to measure infants' attention, memory, speed of processing and representational competence in order to investigate whether a direct link exists with later vocabulary size. They observed that several of these measures predicted later language outcomes. Infants' memory and representational competence were related to language at both 12 and 36 months in a concurrent and predictive way.

Benasich and colleagues (1996, 2002) have also tested cognitive/attentional mechanisms using a visual novelty detection task to understand whether the processing deficit was specific to the auditory domain. Infants were habituated to faces and then tested with a familiar versus a novel face. Atypical infants differed from their typical peers in the rate of habituation as well as in the magnitude of the novelty effect. Moreover, several variables of the rapid auditory processing task and the novelty detection task were correlated. The authors interpreted this as the evidence that the two tasks may be tapping onto similar processes, suggesting that information processing may not be modality specific (Benasich and Tallal, 1996).

2.2 Predictions and objectives

In the light of the above, our study tried to enrich the previous literature addressing five main questions:

1) Are better auditory processors better at learning language?

2) Is there a direct relation between how infants process rapidly presented auditory stimuli (both linguistic and non-linguistic sounds) and later language outcomes?

3) Do auditory processing abilities show important individual differences?

4) Is there a stable link between general cognitive skills and later cognitive and linguistic development?

4) Is it possible to detect early behavioral differences between typical and atypical populations?

In order to address these questions two longitudinal studies were performed (following Benasich and Tallal, 1996, 2002). Each infant performed an auditory discrimination threshold task (using the rapid auditory processing paradigm, as in Benasich and Tallal, 1996), followed by a visual novelty detection task as a control for general cognitive skills.

The auditory discrimination threshold was evaluated using non-linguistic sounds (tones) in one cohort of infants and linguistic sounds (syllables) in another cohort of infants in order to investigate the domain-specificity of the process within the auditory modality. Specifically, this assessed whether differences in the thresholds between linguistic and non-linguistic sounds were observable, as we expected the processing of syllables to be more complex than tones. This could potentially show how the increased complexity of the linguistic domain might affect acoustic processing.

Subsequently, repeated measures of infants' vocabulary (CDIs, Fenson et al., 1993) were taken at 12, 14, 18 and 24 months and a cognitive test, the Mullen Scale of Early Learning (Mullen, 1995), was administered between 18 and 20 months as an additional measure of early processing competences.

Few of the previous studies followed up vocabulary growth systematically and from the earliest age. Thus, it remains unclear whether the predictive effect of the auditory discrimination threshold is stable across language development.

In addition, controlling for external factors is an important challenge in language acquisition research, as they might affect infants' performance in a laboratory environment. Thus, in studies that aim to assess individual variability, taking these factors into account is relevant. The current study has, therefore, used a visual novelty detection task in order to investigate first as previously proposed, whether processing abilities can be considered domain-general (e.g. Tallal et al., 1985a, Benasich & Tallal, 1996). Secondly, it also asked whether measures of cognitive abilities per se could be used as potential predictors of language outcome. Lastly, by administering the Mullen scale at the toddler age, the stability of general cognitive skills across development was investigated.

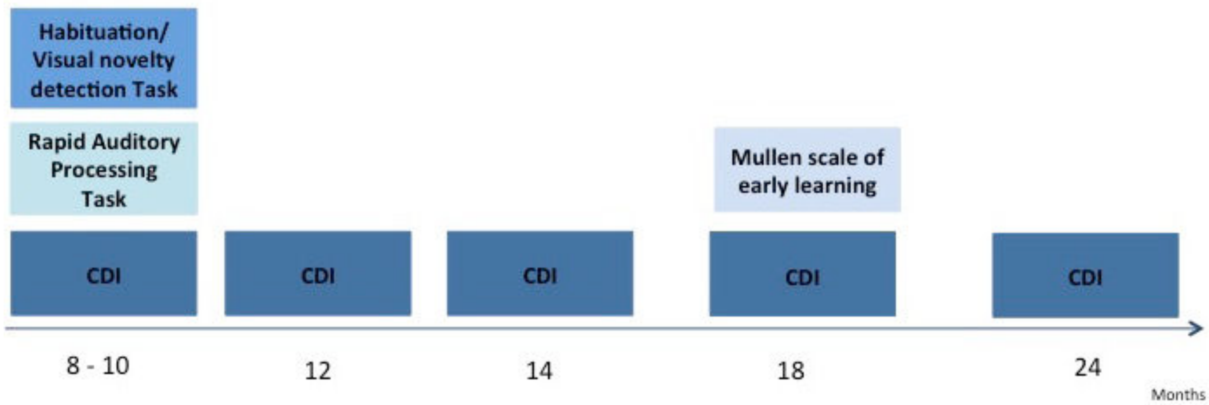
Importantly, our sample consisted of French-learning infants, while previous studies tested English-learning infants. Since the trajectory of vocabulary growth may differ as a function of differences between the grammatical structures of languages (Floccia et al. 2018), it is relevant to test a variety of languages.

Lastly, we explored whether early behavioral markers of risk are detectable at early stages by comparing typical infants and infants at-risk. Moreover, since typical infants may also exhibit relevant individual differences, the lowest 5% of typical participants (in receptive and vocabulary score) was examined to explore individual differences within typical development.

2.3 Study design

At the first visit to the laboratory each infant was tested on the RAP and the habituation/visual novelty detection tasks and parents were asked to fill a short version of the Parental questionnaire of the MacArthur Bates CDI (Fenson et al., 1993) at 9 months. In addition, information about infants' health conditions, socioeconomic status and familial risk of hearing and language learning impairments were collected. All infants were followed longitudinally and parents were asked to complete the online version of the CDI at 12, 14, 18 and 24 months of age. Between 18 and 20 months families were invited for a second visit to the laboratory and a standardized cognitive test, the Mullen Scale of Early Learning (Mullen, 1995), was administered in order to investigate general attention, visual and receptive/expressive abilities. This longitudinal design is similar to those used by Benasich and colleagues (e.g. Benasich et al., 1996, 1998, 2002). The existing differences in the procedures are described in the following sections.

Figure 2.1. *The design of the study*



2.4 Material and methods

2.4.1 Participants

Two cohorts of 9-month-old infants were tested and followed longitudinally with an identical design. In the first cohort (n = 32, 12 girls, mean age: 9 months and 1 day, age range: 8-10 months), the auditory threshold for tone discrimination (non-linguistic stimuli) was investigated, whereas in the second cohort (n = 36, 20 girls, mean age: 9 months and 12 days, age range: 8-10 months), syllables (linguistic stimuli) were presented. Both cohorts were also tested on the same habituation/visual novelty detection tasks in order to explore possible connections between the rapid auditory processing and the habituation tasks and to link cognitive abilities to language development.

A group of infants (n = 11) with a family risk/history of language impairment (developmental dyslexia) was also included (n = 8 in the first cohort, n = 3 in the second cohort).

All methods were approved by the CERES Ethics Committee of the Université Paris Descartes (nr. 2016/32 "Exploring early language acquisition using the Head-turn Preference Procedure"). Parents of all participating infants provided informed consent prior to participation.

For more specific details about the participants in each individual task/measure, see below.

2.4.2 The Conditioned head turn preference procedure (CHTP)

We used a version of the conditioned head turn procedure similar to Benasich and Tallal (1996) with minor changes in the physical setup. Three television monitors were placed inside a sound-attenuated testing room. One monitor was located on the right, one on the left and one in the center (see **Figure 2.2**).

Figure 2.2. *The experimental booth*

On the central screen, a visual attention getter (a video featuring a looming yellow circle) was played. The central screen was used in order to reorient the infant after each head turn. On the side screens, two different animated videos appeared: Big Bird on the left screen and Elmo on the right screen. The videos were used as rewards for the discrimination task.

A video camera placed above the central screen recorded the session and transmitted information, through a monitor, to the experimenter, placed outside the booth and thus blind to the experiment, who controlled the presentation of the stimuli. The sound stimuli were presented through two loudspeakers (on the right and left sides) that were simultaneously activated in order not to provide information about the direction of the sound. The infant was seated on a caregiver's lap, on a chair in the middle of the booth (75 cm away from the central screen). The caregiver listened to masking music and was instructed not to turn his/her head or interact with the infant in order to avoid influencing the infant's response.

The goal of the task was for infants learn to discriminate between two sounds (either tones or syllables) in a pair separated by a silence of varying duration, to associate each sound with a head-direction (left or right), and to make a correct head-turn, if they were able to discriminate the sounds.

2.4.2.1 Experimental procedure

The experiment consisted of three phases: *Shaping*, *Association* and *Variable ISI Test Phase*. The procedure is a partial adaptation of Benasich and Tallal (1996). During the Shaping and Association phases, infants were expected to learn the correct response for each of the two sound sequences: left for Sequence 1 (100Hz_100Hz or da_da) and right for Sequence 2 (100Hz_300Hz or da_ba). Sides were kept the same for all the tested infants.

During Shaping and Association, the tone or syllable pairs were separated by 500 ms ISI (interstimulus interval). In the Variable ISI Test phase, the ISI was gradually decreased from 300 ms to 20 ms by increments of 20 ms upon each correct turn (hit) and increased by the same amount upon an incorrect turn (miss).

I) Shaping:

During the initial 10 trials (5 left, 5 right side randomly presented), reinforcement was not contingent on the correctness of the response. Two seconds after the presentation of the tone or syllable sequence (with ISIs of 500 ms), the appropriate video was automatically displayed on the correct screen (left or right), independently of the infant's response; the video reward lasted 4 s.

II) Association:

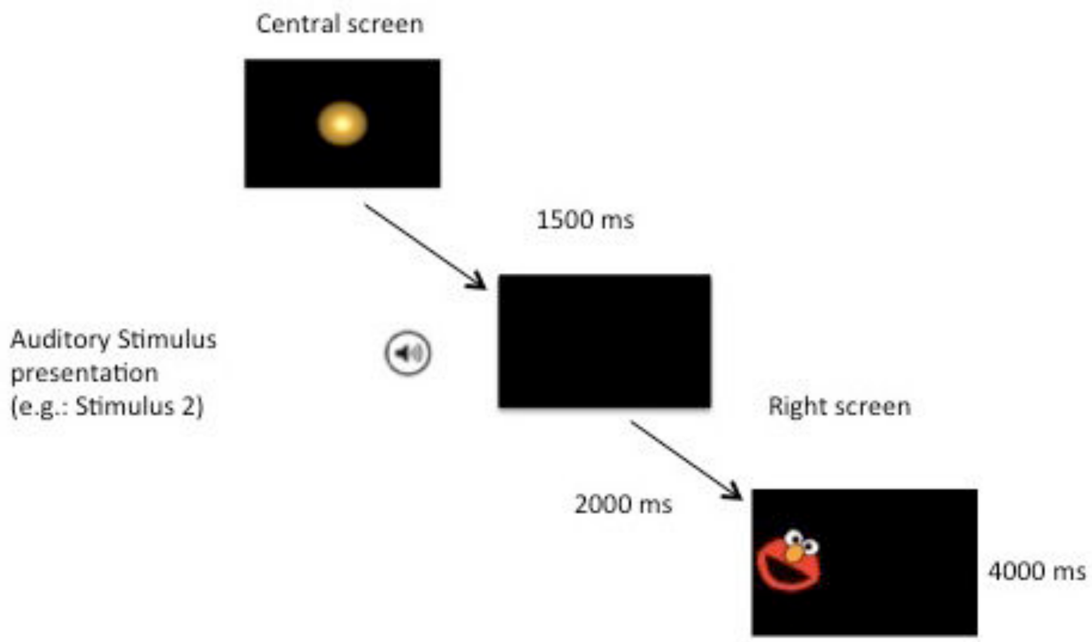
The two Stimuli (1 and 2) were randomly presented with an ISIs of 500 ms, i.e. an ISI value well above infants' expected discrimination threshold. Visual reinforcement here was contingent on a correct response. If the infant made a correct head-turn [left for Sequence 1 (100Hz_100Hz or da_da) and right for Sequence 2 (100Hz_300Hz or da_ba)], the reward video was displayed. Contrarily, after incorrect head-turn, or inability to respond within 4 sec after stimulus presentation, no video reward was displayed and the next trial was presented. To enter the test phase infant has to reach a criterion of six correct responses out of seven consecutive trials. Infants who failed to reach the criterion of 6 consecutive correct trials across the 24 trials of the association did not continue the experiment, as they did not learn the expected responses on which the discrimination test relied.

III) Variable ISI Test Phase

The same contingent response outlined above for the Association phase was followed. Using a "1-up, 1-down adaptive procedure" with a step size of 20 ms, Stimulus 1 and Stimulus 2 were randomly presented with ISIs changing from 300 to 20 ms, until four reversals. A 'reversal' was defined as a shift from a correct to an incorrect response or the opposite (Benasich an Tallal, 1996). The experiment was terminated automatically after four reversals (consecutive or not) were reached.

The ISI was decreased by 20 ms upon every correct head-turn, and increased by the same amount upon every incorrect head-turn, following the adaptive staircase procedure (Levitt 1970). The gradual decrease and increase of the ISI should produce a consistent pattern of "reversals" that converges towards the individual's rapid auditory processing threshold.

Figure 2.3. *Schematic illustration of the experimental procedure*



2.4.2.2 Experiment 1: RAP tones

2.4.2.2.1 Participants

Forty-six 9-month-olds were tested in the tone version of the RAP task using the conditioned head-turn preference procedure. All infants were full-term, had no history of hearing or visual impairment, no recent occurrences of hear infection and no family history of congenital hearing loss. French was the only language spoken in the families.

Among these 46 infants, fourteen were not included in the study because they were not able to reach the test phase (7), they cried or were fussy (3), or they failed to reach the 4 reversal trials during the test phase (4). A final sample of 32 infants (12 girls; mean age: 9 months and 1 day, age range: 8-10 months) participated in the study. Importantly, the 14 infants who did not complete the auditory experiment were not tested in the habituation/visual novelty detection task either.

This choice was made because, following the results of Benasich and Tallal (1996) several variables between the acoustic and the cognitive task showed a correlational pattern. Hence, it was important that each infant performing the acoustic task also be evaluated in his/her more general cognitive abilities.

Moreover, within the group of 32 infants retained for final analysis, eight infants with a documented family history of dyslexia were identified. Information about family risk was assessed through the parental questionnaire administered during the visit. Families reported a history of developmental dyslexia in a first-degree family member (parents and/or siblings) as diagnosed by a speech therapist.

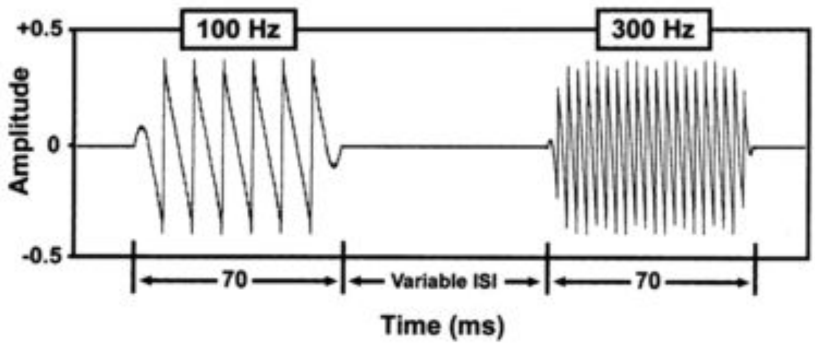
Table 2.1. *n of participants: tones cohort*

	RAP	Habituation /visual novelty detection	CDI_9 months	CDI_12 months	CDI_12 months	CDI_18 months	Mullen scale	CDI_24 months
Total n of participants	32 (F =12)	32 (F = 12)	32 (F = 12)	28 (F = 10)	27 (F = 10)	21 (F = 8)	10 (F = 4)	12 (F = 3)
Typically developing	24	24	24	20	19	14	7	4
At-risk	8	8	8	8	8	7	2	4

2.4.2.2.2 Stimuli

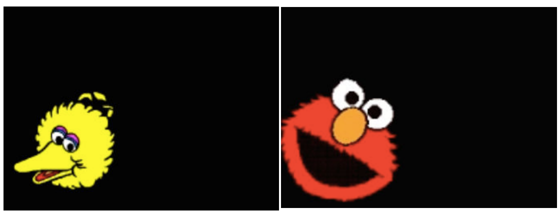
Two tones stimuli were created with duration of 70 ms, rise and fall times of 20 ms and the fundamental frequency of 100 Hz (Tone 1) and 300 Hz (Tone 2). Stimulus sequences were then constructed concatenating Tone 1 followed by Tone 1 (Stimulus sequence 1: 100_100 Hz) and Tone 1 followed by Tone 2 (Stimulus sequence 2: 100_300 Hz). The ISI varied from 500 to 20 ms (the choice of ISI is described in detail below). Both tone sequences were presented at an amplitude of 75 dB, a comfortable conversational sound level (**Figure 2.4**).

Figure 2.4. *Waveform representation of the complex tones used in the RAP task. (from Benasich and Tallal, 2002, p. 37)*



Visual stimuli were a moving cartoon of Sesame Street's Elmo and Big Bird displayed against a black background. These stimuli were adapted from Castellot B., Pons F., Sebastián-Gallés N., Develop. Science. 2011. **Figure 2.5**. Both visual and auditory stimuli were presented through PsyScope, version X B60 run on a Mac OS X, version 10.10.5.

Figure 2.5. *Reinforcement videos displayed, respectively, to the left and right side*



2.4.2.3 Experiment 2: RAP syllables

2.4.2.3.1 Participants

Forty-nine 9-month-olds were tested in the syllable version of the RAP task using the conditioned head-turn preference procedure. All infants were full-term, had no history of hearing or visual impairment, no recent occurrences of hear infection and no family history of congenital hearing loss. French was the only language spoken in the families.

Among these 49 infants, thirteen infants were not included in the study because they did not complete both the RAP and the visual novelty detection tasks. Therefore, 28 infants (16 girls, mean age: 9 months 8 days, range: 8-10 months) were included in the analyses of the RAP task.

Contrarily to the previous cohort, here 8 infants (4 girls) who did not complete the auditory experiment were still tested in the habituation/visual novelty detection task and were included in the cohort. This different choice was made because, after preliminary analysis of the data collected within the tone cohort, no correlation between the RAP and the visual novelty detection tasks has occurred. A final sample of 36 infants (20 girls, mean age: 9 months 12 days, range: 8-10 months) participated in the study.

Within the group, three infants with a documented family history of dyslexia were identified. Information about family risk was assessed through the parental questionnaire administered during the visit. Families reported a history of developmental dyslexia in a first-degree family member (parents and/or siblings) as diagnosed by a speech therapist.

Table 2.2. *n of participants: syllables cohort*

	RAP	Habituation/visual novelty detection	CDI_9 months	CDI_12 months	CDI_14 months	CDI_18 months	Mullen scale
Total n of participants	28 (F = 16)	36 (F = 20)	36 (F = 20)	32 (F = 19)	27 (F = 17)	19 (F = 10)	13 (F = 5)
Typically developing	25	33	33	29	24	17	11
At-risk	3	3	3	3	3	2	2

2.4.2.3.2 Stimuli and procedure

Two syllables, *da* and *ba* were synthesized using a text-to-speech synthesis software (MBROLA) (Dutoit, 1996.) with a pitch of 200Hz (corresponding to the fundamental frequency of female voices) and a phoneme duration of 120ms, resulting in two different stimuli, Stimulus 1: da_da; Stimulus 2: da_ba. The procedure and the video reinforcers were identical to those used in the tone experiment.

2.4.2.4 Data analysis

The experimental software recorded the ISIs used in the test phase for each infant. The output was processed through an in-house script written in Perl in order to determine the psychoauditory thresholds.

Three variables were calculated to evaluate the auditory temporal threshold of each infant: best four reversals, last four reversals and best four succeeded. The average of the ISI tested at the last four reversals is the standard measure in the psychoacoustic literature with adults. However, as infants cannot be given explicit instructions, and thus can comply less well with the task, and tend to have fluctuating attention, this measure may underestimate their real auditory threshold. The two other measures were thus introduced in an attempt to obtain a more realistic estimate of infants' rapid auditory thresholds. Each variable is described in detail below.

1) Best Four Reversals: this variable was calculated as the average of the four best, i.e. lowest ISI values presented at trials that were reversals during test.

2) Last Four Reversals: this variable was calculated as the average of the last four reversals.

3) Best Four Succeeded: this variable was calculated as the average of the four test trials (whether a reversal or not) in which the ISI was lowest and the infant responded correctly.

In addition, three variables were used as a measure of infants' attention/learning in the RAP task: number of trials during the association phase, number of trials during the test phase, number of total trials (association + test phase).

2.4.3 Habituation/visual novelty detection task

This paradigm is widely used as a powerful technique to study discrimination, memory, categorization, discrimination and concept formation in infants. Moreover several studies already investigated its predictive power of the development of general cognitive function during childhood (e.g. Sternberg & Berg, 1985; Bornstein & Sigman, 1986). There are different versions of the procedure that can be applied, here the infant-controlled version, originally developed by Horowitz et al. (1972), was used. This version takes into account that each participant has his/her individual looking time baseline. Changes in looking times during the task are calculated relative to and as a proportion of this initial baseline.

2.4.3.1 Stimuli

A picture of a black and white checkerboard was used as the pre-test image. Infants were habituated to colour images of the same adult face with a neutral expression. Immediately following habituation, each infant was tested with the familiar adult face coupled with the novel child face (**Figure 2.6 A** and **B**). Two test trials were presented. The position and order (left first or right first) of the novel face were counterbalanced between trials and across participants. A video of a turtle was used as attention getter between trials.

Figure 2.6 A. *Visual stimuli used during habituation* **B.** *Visual stimuli used during the test phase*

2.4.3.2 Procedure

Infants performed the task in a sound-attenuated testing room where a central television monitor was placed. Infants were seated on a caregiver's lap, on a chair in the middle of the booth. A video camera placed above the central screen recorded the session and transmitted information through a monitor to the experimenter, placed outside the booth and thus blind to the experiment. The study started with the pre-test trial lasting for a maximum duration of 18s in an infant-controlled manner. During habituation the same adult face appearing simultaneously on the left and the right side of the screen. The same adult face coupled with the face of a young child was used during the test phase as the novel stimulus.
During habituation, the minimum looking time was set to 0.3s (based on Benasich and Tallal, 1996). Looks below this threshold were ignored by the software. Trials ended when a look away larger than 2s occurred. The baseline (100%) looking time was defined as the mean looking time in the first two trials. The same stimuli were repeatedly presented until the habituation criterion was reached. The criterion was set to a looking time equal to or less than

50% of the baseline. The experiment was run with Habit 2.1 on a Mac OS X, version 10.10.5. The session was videotaped, but looking patterns were scored online.

2.4.3.3 Data analysis

Several measures were obtained for each infant on the basis of the online data provided by the experimental software:

The novelty detection measures included (following Benasich and Tallal, 1996):

1) the total looking time calculated as the cumulative time (in seconds) that each infant spent looking at the screen (pre-test + habituation phase + test phase)

2) looking time during the first habituation trial calculated as the cumulative looking time (in seconds) during the first habituation trial

3) the number of habituation trials to criterion (TTC)

4) amount of response decrement (in %) calculated as: [(A – B) / A] X 100]; where A represents the mean of the first two habituation trials and B the mean of the last two habituation trials

5) the linear regression slope (coefficient a) of each infant's looking time across habituation trials

6) novelty preference (in %), calculated as: [N/(F+N) X 100]; where N represents the average looking time for the two test trials (novel items) and F the looking time for the last two trials of the habituation

2.4.4 Standardized questionnaires of vocabulary development

The MacArthur Communicative Development Inventory (CDI, Fenson et al., 1993) is a widely used tool to assess lexical comprehension and production. The inventories also provide a tool for the assessment of communicative gestures and play, early imitation, language comprehension, language production, and the early stages of grammatical development. A primary caregiver reads through a list of words and sentences, and ticks the items that the child understands and/or produces. We used three age-dependent versions of the CDI.

1) At 9 months, during the first visit to the laboratory, the parents of participating infants completed the Inventaire Français du Dêveloppment Communicatif (IFDC), created by Sophie Kern and Frédérique Gayraud (2010). This is a French adaptation of the short version

of the CDI. The parents filled the version for 12-month-olds, as no questionnaire was available for 9-month-olds.

2) Between 12 and 14 months, parents completed the French adaptation of the long online CDI version: 'Words and Gestures' (IFDC version 1999). This version is divided into three sections: first sign of comprehension, receptive and expressive vocabulary (number of words) and gestures.

3) Between 18 and 24 months, parents filled in the French versions of the Hopkins 'CDI: Words and Sentences' and the Hopkins 'CDI: Phrases'.

2.4.4.1 Score calculation

1) At 9 months, the total number of words comprehended by each infant was calculated.

2) At 12 and 14 months, the total number of words comprehended as well as the total number of words produced by each infant was calculated

3) At 18 and 24 months, the total number of words comprehended as well as the total number of words produced by each infant was calculated.

Importantly, the CDI provides gender/age normed language scores assigning infants to percentile ranks ranging from 5 to 99. Crucially, however, as the current study aimed to investigate individual differences, the standardizing procedure was not performed, as it would have caused a significant loss of informative individual data. Thus, the raw score was used by calculating the total number of words comprehended (receptive vocabulary) as well as the total number of words produced by each infant (productive vocabulary).

2.4.5 The Mullen Scale of Early Learning

The Mullen Scale of Early Learning (Mullen, 1995) is an individually administered measure of general cognitive function for infants and pre-schooler children (it can be administered from 0 until 68 months). This test measures different cognitive skills including visual skills, linguistic/auditory skills at the receptive and expressive level, gross and fine motor development. The test has different levels according to the age of the infant/child. Here only the one used in the current study is described.

Infants were tested between 18 and 20 months, thus the stage 5 subtest was used. The stage 5 subtest is intended for infants between 15 months and 0 days and 19 months and 30 days. The test was administered in one of the experimental boxes in which a small table and small chairs were provided in order to place the infant in a confortable environment. Parents were asked to bring one toy from home in order to familiarize with the infant before the beginning of the test. A typical testing session lasted approximately 15 minutes. Parents were allowed to enter the testing room but were instructed not to interact with their infant, unless explicitly requested by the experimenter, as some of the tasks involved a parent-infant interaction.

Five different cognitive skills were investigated: gross and fine motor skills, visual skills, as well as receptive and expressive vocabulary. Each domain included a number of different tasks that the infant had to perform. For each task, a score assessed the infant's performance by assigning one point for each task correctly performed. There was thus a maximum of score that could be achieved in each domain.

1) Gross motor skills: This scale measured central motor control and mobility at 18 months. Motor milestones at this age include walking, running, kicking a ball and completing the tasks in a standing, balanced position. The possible score range is between 0 and 3, 0 meaning that no task has been performed, 3 meaning the accomplishment of all the tasks.

2) Fine motor skills: This scale measured the level of bilateral fine hand movements. The ability to grasp, turn pages and manipulate objects of different sizes is assessed. The possible score range is between 0 and 3 (0 meaning that no task has been performed, 3 meaning the accomplishment of all the tasks).

3) Visual skills: Tasks included discovering masked toys and matching different objects for shape or size. The possible score range is between 0 and 8 (0 meaning that no task has been performed, 8 meaning the accomplishment of all the tasks).

4) Receptive vocabulary: This scale measured the responses to verbal tasks such as pointing to a specific picture or object, understanding requests relative to the spatial environment ("Where is the ball?"), or perform an action ("Can you give me the puppet?"). The possible score range is between 0 (no task performed) and 6 (all tasks performed).

5) Expressive vocabulary: This scale consisted of naming familiar objects and people, combining words with gestures, or by using jabbers and inflections (e.g. by using speech-like sounds modulated in rhythm, containing pauses and ebbs while pointing or manipulating the correct object) if words are non produced. The possible score range is between 0 (no task performed) and 6 (all tasks performed).

The standardized total score is usually calculated by placing each infant in pre-defined levels (percentile). Again in our study row scores have been taken into account for the analyses in order to highlight individual differences.

Figure 2.7. *Picture of a participant performing the Mullen Scale of Early Learning*

2.5 Results

Results of the longitudinal study are discussed here. Two cohorts of infants, one investigating the auditory threshold for tone discrimination and the second for syllable discrimination were tested. The two cohorts performed the same habituation/visual novelty detection task. Moreover, measures of CDIs were collected to explore the predictability of the two tasks on the subsequent language development.

Results of the auditory threshold discrimination are going to be presented separately for the tone and the syllable cohorts. Importantly, following Moore et al., 2008, participants in the two RAP tasks were divided into sub-groups based on their compliance with the task, as indexed by their individual profile of the adaptive staircase procedure. The purpose of this was to distinguish between infants who performed the task well and those whose attention fluctuated during this relatively long and demanding task (see more details below). Results are discussed by subgroups.

In addition, in order to evaluate whether differences in the thresholds between linguistic and non-linguistic sounds occurred, a comparison between the two cohorts is presented as we expected the processing of syllables to be more complex than tones.

Furthermore, data from the habituation/visual novelty detection task as well as significant correlations with language development are discussed together for the two cohorts.

Finally, within the two cohorts, eleven infants with a documented family history of dyslexia were identified. Hence, difference in the behaviour between typical and atypical populations are analysed and discussed separately.

2.5.1 Rapid auditory processing (RAP) results

2.5.1.1 Experiment 1: RAP tones

A total of 32 infants completed the test phase of the RAP task for the tone experiment. As mentioned before, for each infant 3 variables were extracted to measure the auditory temporal threshold: Four best reversals, Four last reversals, Four best succeeded. Moreover, 3 additional measures were included to evaluate attention-performance abilities during the acoustic task: n trials association phase, n trials test phase, total n of trials. **Table 2.3** presents the means and standard deviations of the variables measured in the tone experiment.

Table 2.3. *Means and Standard Deviations of the variables measured in the RAP task (n=32)*

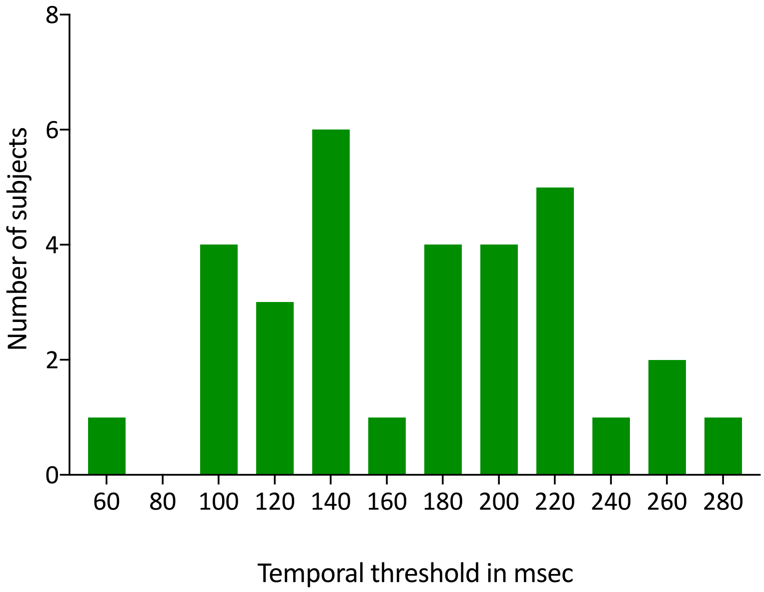
RAP variables	Mean	SD
Four Best Reversals (ms)	169.4	52.7
Four Last Reversals (ms)	181.4	50.3
Four Best Succeeded (ms)	183.4	50.6
n trials association phase	10.9	3.4
n trials test phase	22.6	5.8
Total n of trials	33.5	7.5

Examination of the acoustic variables between males (n= 20) and females (n=12) showed no significant effect of gender. Unpaired t-tests (with Welch's correction that does not assume equal variance) were performed for the variable Four Best Reversals between males, (M = 166.3 s, SD = 60.1) and females (M = 174. 6 s, SD = 39.3), t (29.68) = 0.47, p = 0.64, Ns); for Four Last Reversals between males (M = 175.5 s, SD = 58.6) and females (M = 191.3. s, SD = 31.9), t (29.84) = 0.98, p = 0.33, Ns); and for Four Best Succeeded for males (M = 178.8 s, SD = 58.5) and females (M = 191.3. s, SD = 34.6), t (30) = 0.45, p = 0.45, Ns).

Additionally, as the sample's age ranged between 8 and 10 months, a Pearson product moment correlation was calculated suggesting no significant effect of age for this sample (Four Best Reversal, r = - 0.17, Ns; Four Last Reversals, r = - 0.15, Ns; Four Best Succeeded, r = - 0.21, Ns). The variable n of trials association phase did not show any significant effect of age (r = -0.24) either, suggesting that younger infants did not need more trials in order to reach the test phase.

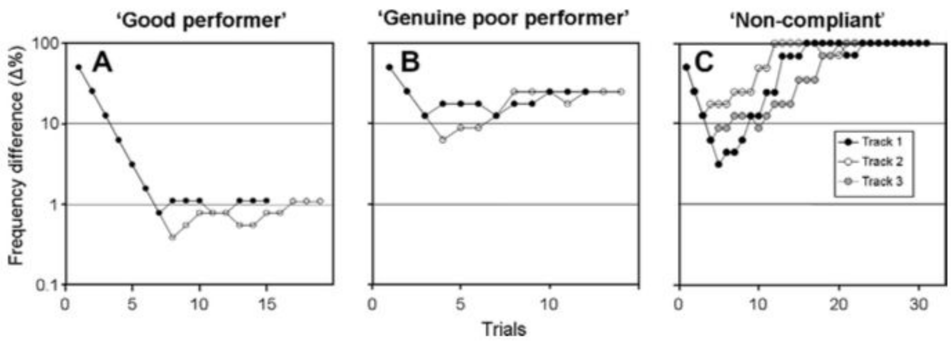
In **Figure 2.8** the histogram of the mean auditory thresholds over the Four Best Reversals is presented. This value was calculated averaging the four best reversals during the test (the lowest ISI) for each participant. Values ranged between a minimum of 60 ms and a maximum 275 ms.

Figure 2.8. *Frequency distribution of the mean auditory temporal thresholds in ms (Four Best Reversals) measured across participants during the test phase.*



Importantly, as expected, infants' performance during the adaptive procedure greatly varied with some infants showing obvious signs of loss of attention. Unlike in many asks, here stimuli are not independent from one another, they are chosen on the basis of performance of previous trials. Loss of attention or failure to understand and follow the task thus have more detrimental effects than in other tasks. However, infants are prone to fluctuations of attention. Following Moore et al. (2008), we thus decided to categorize participants into three different sub-groups as a function of the profile of the adaptive staircase they produced. To analyse data from two frequency discrimination experiments with 6-11-year-old children, Moore et al. (2008) labelled "*good performers*" children who exhibited initial good responses, followed by a rapid convergence towards the auditory threshold, with well-defined reversal points. These children needed few trials to reach the threshold. In the second group, "*genuine poor performers*" showed a similar pattern, but reached a higher absolute threshold. Lastly, "*non compliant*" participants were children who performed well for a certain number of trials, but then their performance declined dramatically (the profile was not staircase-like). The high thresholds obtained in this group are most like due to fluctuations of attention, rather than poor acoustic abilities. **Figure 2.9** provides examples of each profile.

Figure 2.9. *Examples of performance sub-groups from Moore et al., 2008, p. 150*

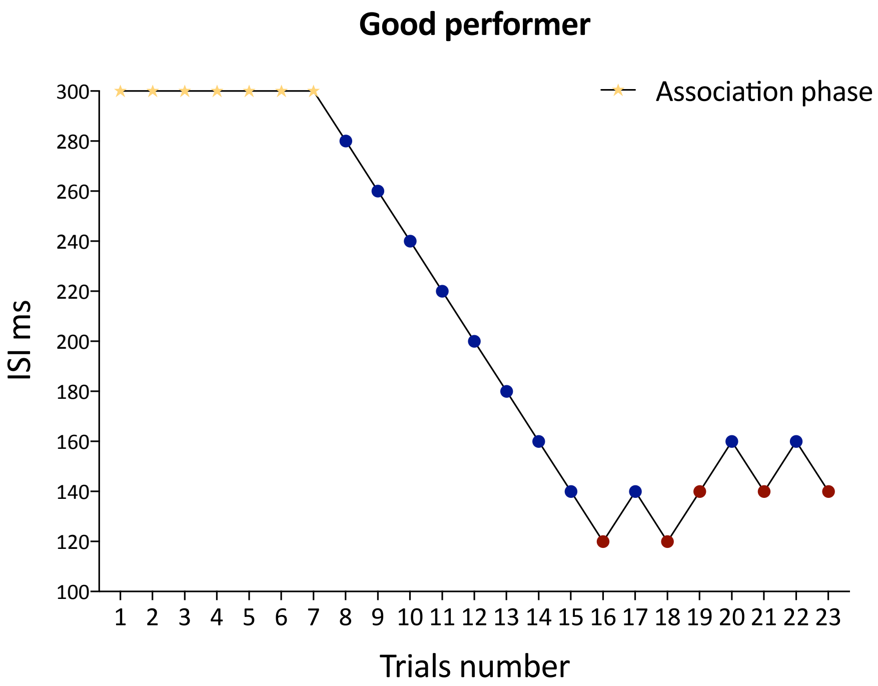


Participants in the tone and syllable cohorts were similarly divided into groups. The *good* and *genuine poor performance* profiles were readily observable in our dataset. But we did not observe *non compliant* patterns. Rather, an *unstable/inconstant performers* group was defined containing participants with erratic profiles. Examples are presented below.

2.5.1.1.1 RAP tones: analysis of the profiles

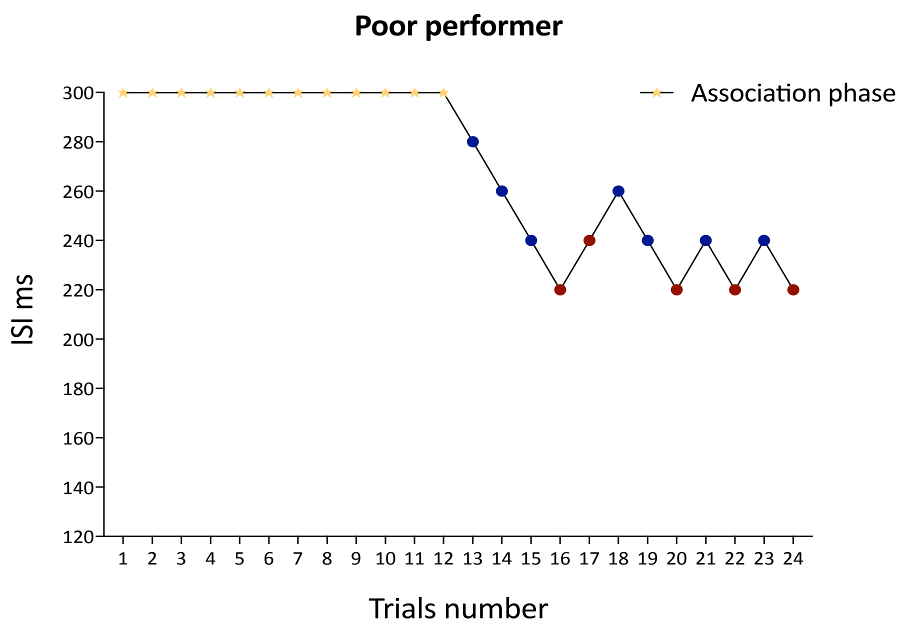
The *good performers* group was the smallest in number (n = 8). A selected participant from the dataset is presented in **Figure 2.10**. Consistent with Moore et al. (2008), participants here exhibited initial good responses followed by a rapid achievement of the auditory threshold and by presenting well-defined reversal points.

Figure 2.10. *Participant included in the good performers profile. Blue dot: hit, red dot: miss.*



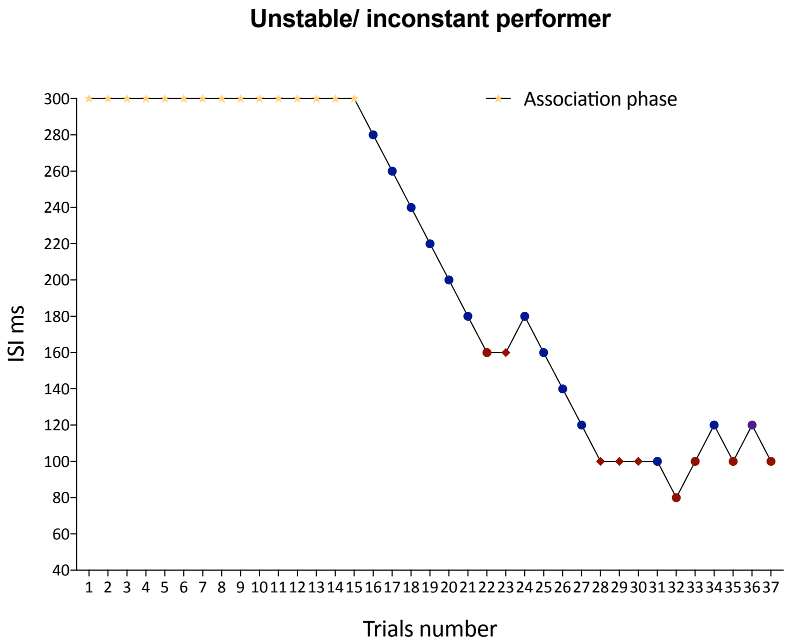
Poor performers (n = 11) were infants who showed a similar pattern as the good ones, but with higher auditory thresholds. In **Figure 2.11** a participant included in this group is presented.

Figure 2.11. *Participant included in the poor performers profile. Blue dot: hit, red dot: miss.*



Lastly, the group of *unstable/inconstant performers* (n = 13) was the biggest in number. Infants in this group presented distant reversal points and several fluctuations in the attentional level during the task (e.g. trials with no responses given). **Figure 2.12** shows one of the participants included in this group.

Figure 2.12. *Participant included in the unstable/ inconstant performers profile. Blue dot: hit, red dot: miss, red rhombus: no response.*



The descriptive statistics of the RAP variables for the three groups is reported in **Table 2.4**. A one-way ANOVA with Factor Group (*good/poor/unstable*) was performed to compare the three acoustic variables: Four best Succeeded: [F (2,29) = 1.287, p = 0.243, Ns.]; Four Last Reversals: [F (2,29) = 3.829, p = 0.003], Scheffe *post hoc* test comparing *poor* and *good* profiles: p = 0.03; Four Best Succeeded: [F (2,29) = 3.63, p = 0.039], Scheffe *post hoc* test comparing *poor* and *good* profiles: p = 0.04. The division in groups based on participants' performance was important in order to analyse performance variability in the acoustic task.

Table 2.4. *Means and (SD) of the variables measured in the RAP task for the three profiles: good performers (n= 8), poor performers (n = 11) and unstable/ inconstant performers (n = 13).*

RAP variables / Profiles	Four Best Reversals (ms)	Four Last Reversals (ms)	Four Best Succeeded (ms)	n trials association phase	n trials test phase	Total n of trials
Good	142.5 (45.9)	147.5 (46)	155 (53)	10.5 (3.7)	20.5 (4.2)	31 (5)
Poor	182.3 (51.4)	207.3 (30.9)	211.8 (36.6)	10.5 (2.7)	21.4 (5.5)	31.8 (7.3)
Unstable	175 (55.6)	180.4 (55.8)	176.9 (50.3)	11.6 (3.9)	24.9 (6.6)	36.5 (8.5)

2.5.1.2 Experiment 2: RAP syllables

A total of 28 infants completed the test phase of the RAP task for the syllables experiment. As before, for each infant the same variables (Four best reversals, Four last reversals, Four best succeeded) were calculated to evaluate the auditory temporal threshold as well as measure of attention/ performance abilities during the acoustic task (n trials association phase, n trials test phase, total n of trials). **Table 2.5** shows means and standard deviations for all variables.

Table 2.5. *Means and Standard Deviations of the variables measured in the RAP task (n=28)*

RAP variables	Mean	SD
Four Best Reversal (ms)	188.6	47.4
Four Last Reversal (ms)	200.7	41.4
Four Best Succeeded (ms)	202	36.5
n trials association	10.2	3.2
n trials test phase	22.6	4.8
Total n of trials	32.8	7.4

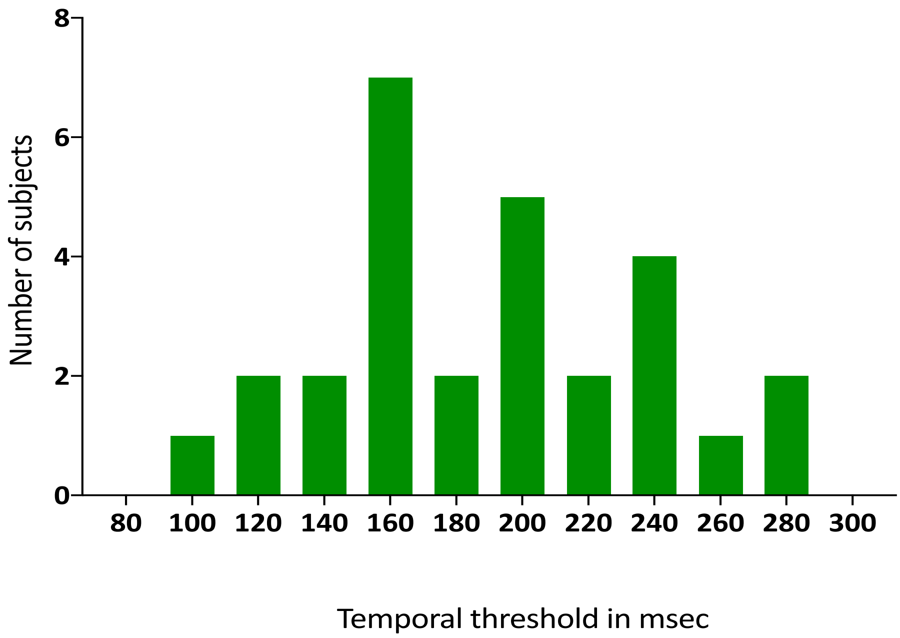
Examination of the acoustic variables between males (n = 12) and females (n =16) did not show any significant effect of gender. Unpaired t-tests (with Welch's correction that does not assume equal variance) performed for the variable Four Best Reversals between males, (M = 188.3 s, SD = 58.5), and females (M = 188.8 s, SD = 33.6), t (18.32) = 0.02, p = 0.98, Ns); Four Last Reversals between males (M = 200.4 s, SD = 51.6) and females (M = 200.9. s, SD

= 33.6), t (17.33) = 0.03, p = 0.97, Ns) and Four Best Succeeded between males (M = 200.4 s, SD = 46.1) and females (M = 203.1 s, SD = 28.9), t (17.33) = 0.001, p = 0.86, Ns).

Additionally, as the sample's age ranged between 8 and 10 months, a Pearson product moment correlation was calculated suggesting no significant effect of age for this sample (four best reversal, r = 0.31, Ns; four last reversals, r = 0.28, Ns; four best succeeded, r = 0.08, Ns; n of trials association phase, r = 0.14). Consistently with the tone experiment, younger infants did not need more trials in order to reach the test phase.

The frequency distribution of the mean auditory thresholds over the last Four Best Reversals is presented in **Figure 2.13**. The values ranged between a minimum of 90 ms and a maximum 285 ms.

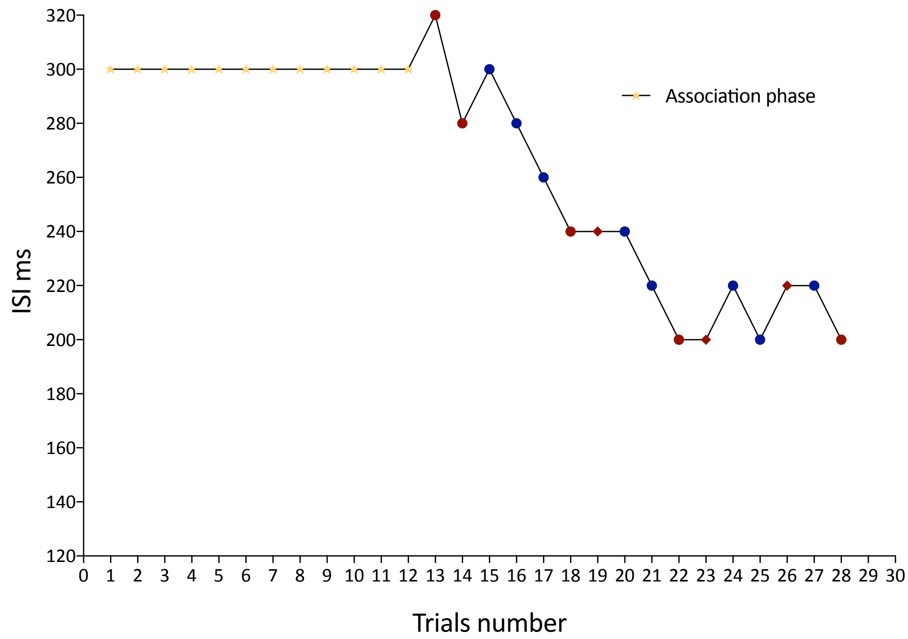
Figure 2.13. *Frequency distribution of the mean auditory temporal thresholds in ms (Four Best Reversals) measured across participants during the test phase.*



2.5.1.2.1 RAP syllables: analysis of the profiles

We enter the participants into the same three sub-groups as above: the good performers (n = 11), the poor performers (n = 8) and the unstable/inconstant performers (n = 9). An example of an *unstable/inconsistent* profile is presented in **Figure 2.14**.

Figure 2.14. *Participant included in the unstable/inconstant performers profile (syllables cohort). Blue dots: hit, red dots: miss, red rhombus: no response.*



The descriptive statistics of the RAP variables found in the three groups is reported in **Table 2.6**. A one-way ANOVA with Factor Group (*good/poor/unstable*) was performed to compare the three acoustic variables: Four best Reversals: [F (2,25) = 5.657, p = 0.009]; Scheffe *post hoc* test comparing *good* and *unstable* profiles: p = 0.01; Four Last Reversals: {F (2,25) = 7.727, p = 0.002}, Scheffe *post hoc* test comparing *good* and *unstable* profiles: p = 0.002; Four Best Succeeded: {F (2,25) = 7.794, p = 0.002}, Scheffe *post hoc* test comparing *good* and *unstable* profiles: p = 0.003. Contrarily to the tones, between the groups of *good* and *unstable* performers several comparisons reached the level of significance.

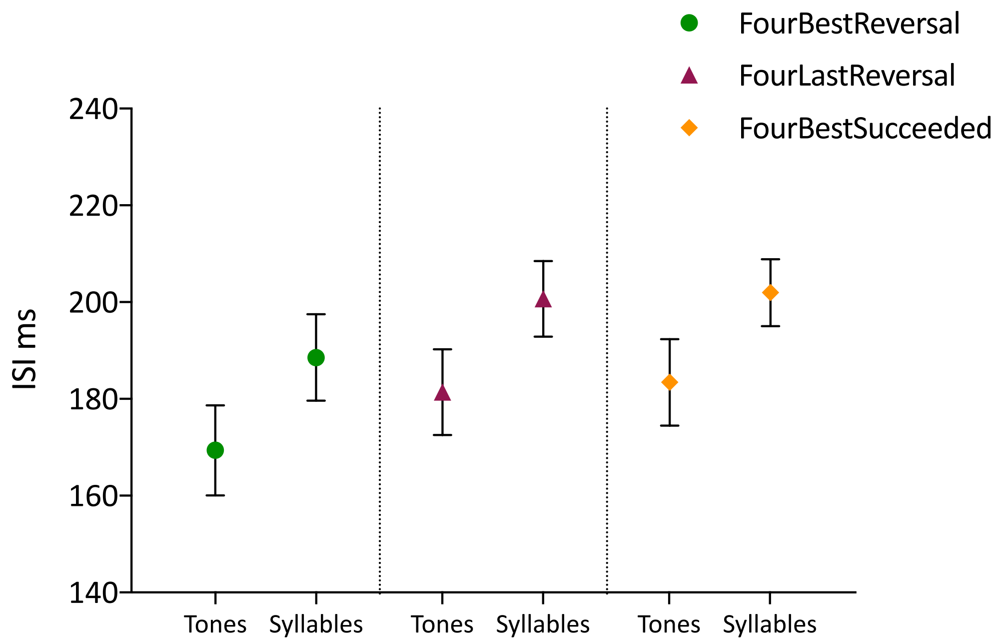
Table 2.6. *Means and (SD) of the variables measured in the RAP task for the three profiles: Good performers (n= 11), Poor performers (n = 8) and unstable/ inconstant performers (n = 9).*

RAP variables / Profiles	Four Best Reversals (ms)	Four Last Reversals (ms)	Four Best Succeeded (ms)	n trials association phase	n trials test phase	Total n of trials
Good	160 (38.7)	173.6 (32.3)	175.9 (27.2)	8.8 (3)	22.4 (5.6)	31.3 (8.1)
Poor	190.6 (26.5)	201.3 (23.7)	208.8 (22)	11.4 (2.8)	23 (4.4.)	34.4 (6.7)
Unstable	221.7 (53.2)	233.3 (42.1)	227.8 (37.8)	10.8 (3.4)	22.6 (4.6)	33.3 (7.5)

2.5.1.3 RAP: Comparison between syllable and tone discrimination

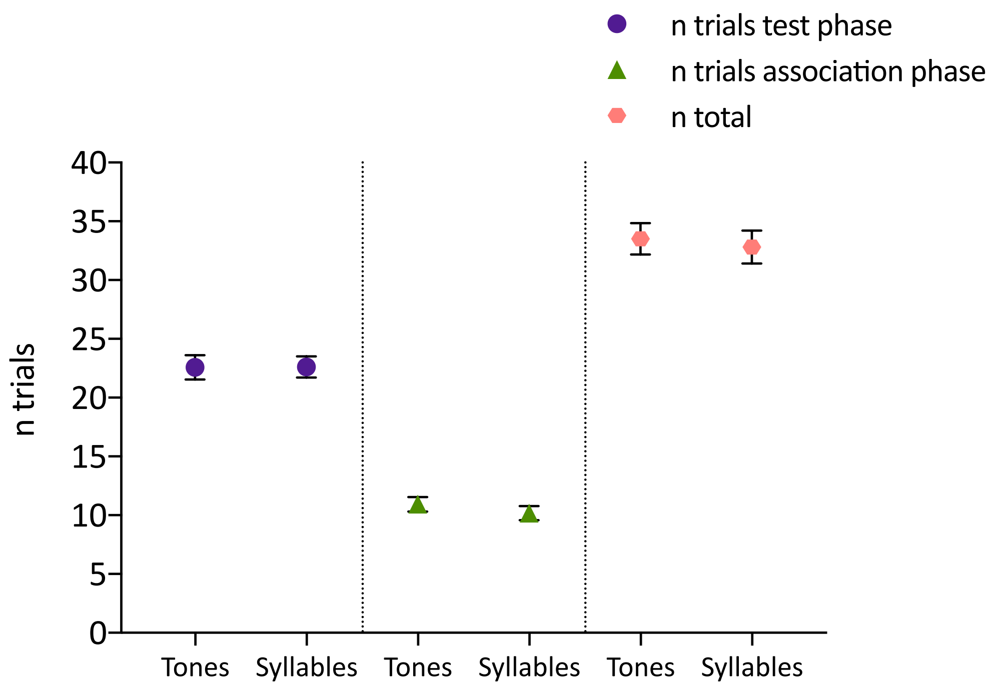
To directly compare the tasks, unpaired t-tests (with equal variance not assumed) were performed. Contrarily to the initial expectation, being exposed to linguistic or non-linguistic sounds did not affect the auditory thresholds: Four best reversals: tones (M = 169.4, SD = 52.7) vs. syllables (M = 188.6, SD = 47.4), t (57.95) = 1.48, p = 0.1, Ns; Four last reversals: tones (M = 181.4, 50.3) vs. syllables (M = 200.7, SD = 41.4), t (57.8) = 1.62, p = 0.1, Ns; Four best succeeded: tones (M = 183.4, SD = 50.6) vs. syllables (M = 202, SD = 36.5), t (56.09) = 1.64, p = 0.1, Ns) In **Figure 2.15** these values are report graphically.

Figure 2.15. *Comparison of the three acoustic variables measured in the tone (n = 32) and syllable (n = 28) cohorts. Error bars represent the s.e. of the mean.*



Moreover, no significant differences were found comparing how fast infants reached the test phase (n trials association phase): tones (M = 10.9, SD = 3.4) and syllables: (M = 10.2, SD = 3.2) and for the n of trials performed during the test phase: tones (M = 22.6, SD = 5.8) and syllables (M = 22.6, SD = 4.8). Consequently, the total number of trials performed was similar: tones (M = 33.5, SD = 7.5) and syllables (M = 32.8, SD = 7.4) (cf. **Figure 2.16**).

Figure 2.16. *Comparison of the three-attention/ performance variables measured in the tone (n = 32) and syllable (n = 28) cohorts. Error bars represent the s.e. of the mean.*



Additionally, a comparison between types of performance (good, poor and unstable performers) was run between the two cohorts.

Within each group, the obtained auditory thresholds were compared between the tones and the syllables experiments. Coherently with the differences found between the unstable/ inconstant profiles for syllables and tones, only two variables reached the level of significance (**Table 2.7**): (t-tests) Four Last Reversals: *unstable performers tones*: M = 180.4, SD = 55.8; *syllables*: M = 223.4, SD = 43.1; t (19.79) = 2.53, p = 0.02; Four Best Succeeded: *unstable performers tones*: M = 179.6, SD = 50.3; *syllables*: M = 227.8, SD = 37.8; t (19.81) = 2.706, p = 0.01) (unpaired t test with equal variance not assumed).

Table 2.7. *Means and Standard Deviations (SD) of the variables measured between unstable performers in the tone (n= 13) and syllable (n = 9) experiments (* p < 0.05, two-tailed)*

RAP variables	Unstable performers: tones (n=13)	Unstable performers: syllables (n=9)
Four Last Reversals (ms)	180.4 (55.8)	223.3 (43.1)*
Four Best Succeeded (ms)	176.9 (50.3)	227.8 (37.8)*

Overall, the measured auditory thresholds were high in absolute values in both cohorts. The best auditory thresholds measured were about 170 ms for tones and about 180 ms for syllables. These values are much higher than those previously registered (e.g. about 70 ms, form Benasich ant Tallal, 1996). This difference in absolute thresholds may be due to methodological limitations., These limitations will be discussed in section below, after the results of the visual novelty detection task

2.5.2 Habituation/ visual novelty detection task

2.5.2.1 Tone and Syllable cohorts

From the tone cohort, 32 infants were included in the analysis. From the syllable cohort, 28 infants that were included in the RAP also performed the Habituation/VRM task. In addition, 8 infants who did not perform the RAP task did the Habituation/VRM task and were included in the analysis and followed longitudinally. Thus a final sample of 36 subjects was included in the analysis.

The variables of interest were: the number of trials to criterion (TTC) and the linear regression functions (coefficient α) of infant's looking time (in s) across habituation trials. Moreover, measures of individual looking time were calculated for each infant (first looking length and total looking time). The difference between the first two looks during the habituation phase and the two during the test phase was calculated as a measure of the response decrement (in %). Means and standard deviations for all the variables are presented below (**Table 2.8**).

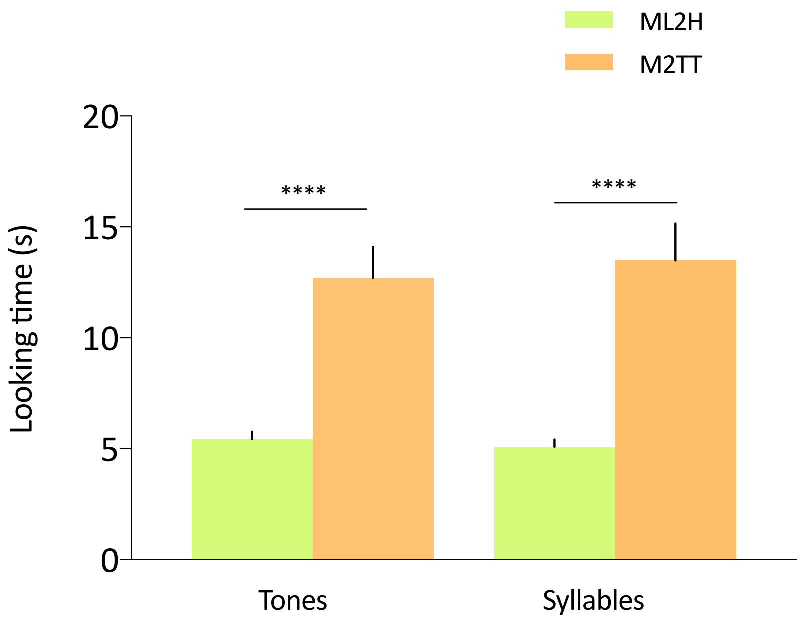
The % of response decrement was the only variable showing a difference between the two cohorts (unpaired t-test with equal variance not assumed; tones: M = 35.2%, SD = 35.8, syllables: M = 56.5%, SD = 12.5, $t (37.67) = 3.19$, $p = 0.002$). This difference was most likely a spurious result, as the overall performance of the two cohorts was similar.

Table 2.8. *Means and Standard Deviations of the variables measured in the Habituation/visual novelty detection task for the tone (n=32) and syllable (n = 36) cohorts (**p < 0.005, two-tailed)*

	Tone Cohort (n = 32)		Syllable cohort (n = 36)	
Habituation/ visual novelty detection variables	**Mean**	**SD**	**Mean**	**SD**
First looking length (s)	12.8	7.4	13.9	7.4
Total looking time (s)	79.4	2.6	88.8	37.2
TTC	5.7	2.6	6,3	1.8
Habituation slope (α)	-0.3	0.3	-0,4	0,4
% Novelty effect	66.7	12	70.0	8.9
% Response Decrement	35.2	35.8	56.5**	12.5

Infants' novelty preference (in %) was used as a measure of visual novelty detection, showing that it significantly exceeded chance (50%) both in the *tone* (M = 66.7, SD = 12; $t (31) = 7.86$, $p < 0.0001$) and the *syllable* (M = 70.0, SD = 8.9; $t (35) = 13.54$, $p < 0.0001$) cohorts. It shows that infants successfully detected the novel face. Difference in the mean looking time of the two last habituation trials (ML2H) and the two trials during the test phase (M2TT) was significantly different in both groups (**Figure 2.17**). Paired t-test was run for the *tones* cohort: ML2H = 5.45 s, SD = 1.89; M2TT = 12.72; SD = 7.09; t (31) = 5. 27, $p < 0.0001$; as well as for the *syllables* cohort: ML2H = 5.08 s; SD = 2.96; M2TT = 13.48; SD = 10.04; $t (35) = 5.51$, $p < 0.001$.

Figure 2.17. *Comparison of the MLH2 and M2TT for the tone and syllables cohort (X-axis). The Y-axis shows the looking time in seconds. Error bars represent the s.e. of the mean*



Relevantly, no effect of age was reported for any of the variables (e.g. *syllables*: Pearson's correlation % novelty effect/age: r = -0.12; *tones*: % novelty effect/age: r = -0.13).

Moreover, none of the reported measured showed a significant effect of gender between males (n = 20) and females (n =12) in either cohort: females (n= 20), males (n = 16). Data are reported in **Table 2.9**.

Table 2.9. *Means and (Standard Deviations) of the Habituation/visual novelty detection variables measured between males and females in the two cohorts.*

	Tones		**Syllables**	
Habituation/ visual novelty detection variables	Males	Females	Males	Females
First looking length (s)	12 (7.7)	14.3 (6.9)	13 (4.9)	14.6 (9)
Total looking time (s)	77.4 (31.3)	82.8 (46.8)	84 (36.1)	89.7 (39.4)
TTC	5.7 (2.9)	5.8 (2.1)	6.1 (1.6)	6.5 (2.1)
Habituation slope (α)	-0.3 (0.4)	-0.3 (0.1)	-0.5 (0.3)	-0.4 (0.4)
% Novelty effect	67.4 (11.1)	65.6 (13.8)	72.4 (10.6)	68.1 (6.9)
% Response Decrement	27.2 (40.1)	48.6 (22.7)	57.5 (15.1)	53.9 (13.5)

Overall, it seems that being exposed to tone or syllable stimuli during the acoustic task did not impact the later performance in the habituation/ visual novelty detection task. Thus, in line with the previous comparison, participants in the two cohorts seemed to behave in a similar manner.

Importantly, previous evidence reported correlations between several variables of the rapid auditory processing task and the novelty detection task suggesting that information processing may not be modality specific (Benasich and Tallal, 1996). Therefore, correlations were run for both cohorts (syllable and tone) between all the variable measured during the RAP task (Four Last Reversals, Four Best Reversals, Four Best Succeeded, n trials association phase, n trials test phase, total n of trials) and those measured during the habituation/ visual novelty detection task (first looking length, total looking time, TTC, slope (α), % novelty effect and % response decrement). Nevertheless, no significant correlation was found between any of the variables tested. Possible interpretations are discussed below.

2.5.3 Discussion: rapid auditory processing and habituation/visual novelty detection tasks

The first phase of the longitudinal study evaluated infants' auditory temporal threshold using the Rapid Auditory Processing Paradigm (RAP) together with a measure of general cognitive abilities using a habituation/visual novelty detection task. The study replicated the design of Benasich and Tallal (1996).

In two cohorts of infants, the discrimination thresholds for linguistic (syllables) and non-linguistic sounds (tones) were established using the same paradigm. This comparison was important to investigate the language-specificity of the process within the auditory modality.

After examination of the results, it seems that the auditory thresholds measured here were not fully representative of infants' real discrimination abilities. The mean auditory thresholds were higher in absolute value than previously reported (e.g. Trehub et al., 1995, Benasich and Tallal, 1996). Complexities in the experimental design might have interfered with a correct performance of the task (for further discussion see section below: limitations of the study).

In addition to the specific methodological limitations encountered, there is already considerable literature reporting a consistent data loss (high attrition rate) in infants' psychoacoustic studies (e.g. Berg & Smith, 1983; Aslin et al., 1983). The 30.4% and the 42.8 % of the tested infants were excluded from final data analysis in the tone and syllable cohorts, respectively. The higher % of infants excluded in the syllable cohort goes in the direction of the original prediction that auditory thresholds for linguistic sounds may be worse than for tones.

It is important to note that high individual variability was observed. In addition to the difficulty of measuring the acoustic ability per se, several other factors (e.g. attention, stimuli, specific procedure) might influence whether participants perform the task successfully. To take this into account, participants were divided into three groups on the basis of the shape of the adaptive procedure during the test phase. Thresholds of good performers were lower than the threshold of the other infants, but they were still relatively high compared to other values reported in the literature.

In addition to auditory discrimination, the study also investigated general cognitive abilities and their relation to auditory thresholds at this early stage of development. In the study performed by Benasich and Tallal (1996) authors found that different measures of speed of encoding during the habituation/visual detection task were significantly correlated with measures of auditory threshold, suggesting the possibility of a close relation between visual and acoustic processing. Contrarily to previous data, no correlations were found between psychoacoustic and visual-detection measures. Most probably, it was the inaccurate evaluation of the temporal thresholds in the current task that has led to this null result.

Lastly, no significant differences between the habituation/visual novelty detection abilities were found between the two cohorts meaning that, being exposed to syllables or tones during the acoustic procedure did not influence the performance in the visual detection task.

Overall, the insensitivity of the auditory threshold task makes it hard to assess auditory processing efficiency in the current study.

2.5.3.1 Auditory threshold discrimination: limitations of the study

In order to correctly interpret the results of the auditory measurements, in this section possible methodological limitations are discussed. As extensively reported, the mean auditory thresholds measured for both linguistic (180ms) and non-linguistic (160ms) sounds were overall higher than previously obtained (e.g. about 70 ms in a similar task in Benasich and Tallal, 1996; 11ms in a gap detection task in Trehub et al. 1995).

It is not uncommon in the infant and child psychoacoustic literature to find widely varying absolute thresholds for the same discrimination tasks. First, it seems that these values are strongly dependent on the specific paradigm and stimuli that are tested. For instance, in a frequency discrimination task, one study reported a mean auditory threshold of 214Hz at 6 years of age (Bishop & Sutcliffe, 2005), while other studies reported thresholds as low as 20-40Hz in infants (Werner Olsho et al. 1982). This confirms the idea that there are low-level methodological differences that influenced psychoacoustic measurement. To date, it is not clear which paradigm is most successful in young infants.

This study was created following the longitudinal design of Benasich & Tallal (1996). However, there were some differences in the procedures. In the original design, during the test phase, probe stimuli (500 ms ISI) occurred after two no/missed responses trials. Inability to respond to two consecutive probe stimuli automatically ended the test phase. The authors made this choice to reduce the impact of inattention on the results (from Benasich and Tallal, 1996). However, this also lengthens the duration of the experiment, so we decided not to follow this. In the current design, we used no probe stimuli. Even so, the mean duration of the experiment was about 7-8 minutes. In order to provide valid threshold data, each infant had to complete three different phases (shaping-association and test phase). Furthermore, the most difficult discriminations, when the ISIs were shorter, occurred close to the end of the procedure. Therefore, several lapses in attention likely occurred.

Some difference in the physical setup also existed. We used a computer screen, while the original study used a different display with a central red flickering light and two dark Plexiglas boxes placed to the left and the right that were moving electrical toy were activated as a reward for a correct response.

Overall, consistently with the high variability in the literature on infants' auditory thresholds, our paradigm might not have been sufficiently sensitive. Even if this specific set of data could be considered only partially reliable and not applicable to predict later language outcomes, they contribute to the literature by highlighting limitations of the methods.

2.5.4 Typically developing infants vs. infants at risk

In order to investigate whether specific differences between typical and atypical populations occurred and predicted later language outcomes, the eleven infants at-risk were analysed separately. Importantly, 8 at-risk infants performed the threshold discrimination task for tone stimuli and 3 for syllables. In the tone cohort, we thus compared the auditory performance of these infants with their typical peers. Due to the small sample size of at-risk infants in the syllable cohort, this comparison was not made.

In line with previous studies (e.g. Benasich and Tallal, 1996), we expected the atypical group to exhibit a poorer auditory threshold and lower general cognitive abilities than the typically developing infants.

As the habituation task was the same between the two cohorts, results between the tone and the syllables cohorts were pooled together both for the typical (n = 57) and the atypical (n = 11) participants.

2.5.4.1 Auditory thresholds: typically developing infants vs. infants at risk (tone cohort)

The 24 typically developing infants were compared with the at-risk group (n = 8). No differences were found (descriptive statistic in **Table 2.10**). Surprisingly, at the group level, infants at-risk exhibited numerically, although not statistically, lower auditory thresholds than typical ones. This result was not in line with the previous literature reporting that infants at risk performed worst than peers (e.g. Benasich and Tallal, 1996; Tallal et al., 1985). As the auditory thresholds were not measured accurately, no conclusions can be drawn from this null result.

However, crucially, infants at risk needed significantly more trials to reach the test phase: typically developing (n = 24): M = 9.8, SD = 2.8; at-risk (n = 8): M = 14.25, SD = 3.1; t (11.21) = 3.608, p = 0.004. This measure is particularly relevant since, performing more trials in this phase reflected slower and poorer discrimination abilities.

Table 2.10. *Means and Standard Deviations of the variables measured in the RAP task for the typically developing (n=24) and at-risk (n = 8) infants in the tones cohort (* p < 0.05, two-tailed, **p < 0.005)*

	Typically developing (n=24)		**At-risk (n = 8)**	
RAP variables	Mean	SD	Mean	SD
Four Best Reversal (ms)	173.1	48.6	158.1	66.1
Four Last Reversal (ms)	187.9	43.5	161.9	66.2
Four Best Succeeded (ms)	192.7	42.6	155.6	64.8
n trials association	9.8	2.8	14.25**	3.1
n trials test phase	21.3	4.3	26.4	8.3
Total n of trials	31.2	5.3	40.6*	9.1

In addition, eight infants from the group of typically developing participant were randomly selected and directly compared with the eight infants at-risk to equate for sample size. Descriptive statistics, reported in **Table 2.11**, confirmed the results obtained for the whole group. No differences were found in the auditory threshold. Nevertheless, the significance in the n of trials to reach the test phase was maintained (unpaired t tests with equal variance not assumed: typically developing (n = 8) M = 8, SD = 1.9; at-risk (n = 8): M = 14.25, SD = 3.1; $t\ (11.8) = 4.88$, $p = 0.0004$.) Moreover, the difference in the total n of trials was also significant: typically developing (n = 8) M = 27.5, SD = 4.5; at-risk (n = 8): M = 40.6, SD = 9.1; $t\ (10.21) = 3.67$, $p = 0.004$).

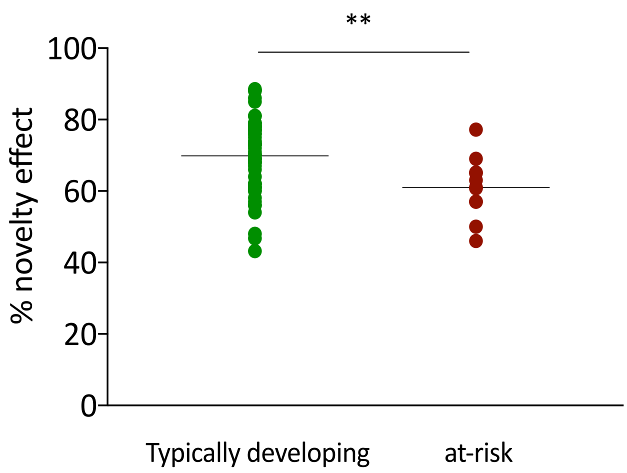
Table 2.11. *Means and Standard Deviations of the variables variable measured in the RAP task between 8 randomly selected typically developing and at-risk infants (n = 8) in the tones cohort (**p < 0.005, two-tailed, *** p < 0.001)*

	Typically developing (n=8)		**At-risk (n = 8)**	
RAP variables	Mean	SD	Mean	SD
Four Best Reversal (ms)	154.3	43	158.1	66.1
Four Last Reversal (ms)	189.3	26.9	161.9	66.2
Four Best Succeeded (ms)	195	30.8	155.6	64.8
n trials association	8	1.9	14.25***	3.1
n trials test phase	19.5	2.8	26.4	8.3
Total n of trials	27.5	4.5	40.6**	9.1

2.5.4.2 Habituation/ visual novelty detection task: typically developing infants vs. infants at risk (tone and syllable cohorts)

Examination of atypical infants' novelty preference revealed that the mean novelty effect (%) significantly exceeded chance (50%) (M = 61.92, SD = 8.61, $t\ (7) = 3.09$, $p = 0.01$). Nevertheless, the % of novelty effect in this population was lower than in the typical group (n = 57), M = 69.9, SD = 10.3; at-risk: (n = 11): M = 61.02, SD = 8.61; $t\ (16.05) = 3.017$, $p = 0.008$ (**Figure 2.18**).

Figure 2.18. *Plot of the individual variability of the % of novelty effect between typically developing (n=57) and infants at-risk (n=11)*



This finding was in line with previous evidence showing that the % of novelty effect is a sensitive predictor of cognitive and linguistic outcomes for both typical and atypical populations (e.g. Rose & Feldman, 1995). Importantly, this result might contribute to the understanding of whether relations between perceptual and cognitive measures in infancy occur and whether they can predict later language and cognitive achievements.

Among the habituation variables, only the habituation slope (α coefficient, $p = 0.03$) differed significantly between the two groups (further details in **Table 2.12**).

Table 2.12. *Means and Standard Deviations of the variables measured in the Habituation/visual novelty detection task for the typically developing (n=57) and at-risk (n=11) infants in the two cohorts (* $p < 0.05$, two-tailed, **$p < 0.005$)*

	Typically developing (n = 57)		**At-risk (n = 11)**	
Habituation/ visual novelty detection variables	Mean	SD	Mean	SD
First looking length (s)	13	7.2	15.68	8.2
Total looking time (s)	86	38.8	75.8	27.79
TTC	6	2.3	6	1.84
Habituation slope (α)	-0.4	0.4	-0.25*	0.13
% Novelty effect	69.9	10.3	61.02**	8.61
% Response Decrement	45.2	28.6	53.07	25.42

As previously, eleven infants from the typically developing group were randomly selected and compared to the infants at-risk to equate for sample size (data in **Table 2.13**). The difference in the % of novelty effect was maintained: typically developing (n = 11) M = 72.3, SD = 5.32; at-risk (n = 11): M = 61.02, SD = 8.61; t (16.66) = 3.68, p = 0.001. None of the other variables, except for the first looking length (p = 0.04), reached significance.

Table 2.13. *Means and Standard Deviations of the variables measured in the Habituation/visual novelty detection task between 11 randomly selected typically developing and at-risk infants in the two cohorts (* p < 0.05, two-tailed, **p < 0.005)*

	Typically developing (n = 11)		**At-risk (n = 11)**	
Habituation/ visual novelty detection variables	Mean	SD	Mean	SD
First looking length (s)	9.7	3.61	15.68*	8.2
Total looking time (s)	81.1	31.8	75.8	27.79
TTC	6.4	2.5	6	1.84
Habituation slope (α)	-0.4	0.17	-0.25	0.13
% Novelty effect	72.3	5.32	61.02**	8.61
% Response Decrement	46.6	22.6	53.07	25.42

2.5.4.3 Discussion: typically developing infants vs. infants at risk

This comparison evaluated the performance of the auditory temporal threshold and the novelty detection performance of infants considered at risk for developmental dyslexia and compared them to typically developing ones. Although the atypical sample was not particularly large, some relevant differences were observed in both tasks.

As discussed before, the results of the acoustic measures cannot be considered fully reliable. For this reason the comparison between the two populations did not converge with the previous literature, where infants at risk exhibited higher auditory thresholds (e.g. Benasich & Tallal, 1996). Nevertheless, the thresholds measured for the atypical group (158ms) were closer to those previously reported for the at-risk population (148ms) in Benasich and Tallal (1996).

However, relevantly, infants at risk needed significantly more trials to reach the test phase. To recall, infants needed a criterion of six consecutive correct trials in the association phase to reach the test. Therefore, performing more trials during this phase perhaps reflected either poorer/slower discrimination abilities or a higher number of inattentive trials (e.g. null responses). Slower processing might interfere with extracting sequential information both in the auditory and the visual modalities.

Looking behaviour in the habituation/ visual novelty detection task also reflected some differences between the two populations. Importantly, the % of novelty effect, reflecting the accuracy and the ability to distinguish familiar from novel stimuli, was lower in the atypical group. Even if at the group-level, they performed differently from chance, the performance was lower than in the typical group. One of the possible interpretations is that infants exhibiting limitations in recognition and memory are also slower or less efficient in retrieving relevant information.

We now turn to the question of whether these results and the relative differences in the two populations serve as predictors of later language and cognitive outcomes, as assessed by expressive and receptive language (CDI measures) as well as general cognitive performance (Mullen Scale of Early Learning).

2.5.5 Standardized questionnaires of vocabulary development

Raw scores were calculated for the cognitive and language measures administered at 9, 12, 14, 18 and 24 months. At 9 months, the CDI data showed a floor effect, and will thus not be presented.

Correlations between RAP and the visual novelty detection variables, on the one hand, and subsequent language measures, on the other hand, were calculated in both cohorts. For the tone cohort, data is available at 12, 14, 18 and 24 months. For the syllable cohort, data is available at 12, 14 and 18 months, since infants participating in this group are, to date, younger than 24 months. Correlations between the CDI scores and all RAP and habituation/visual novelty dependent variables were calculated. The only variable that showed a significant correlation was % novelty effect. Thus only this one is reported below.

Moreover, differences in the receptive and expressive vocabularies of typical and atypical infants were also investigated, as well as the relative patterns of correlations.

Finally, data from the Mullen Scales are reported.

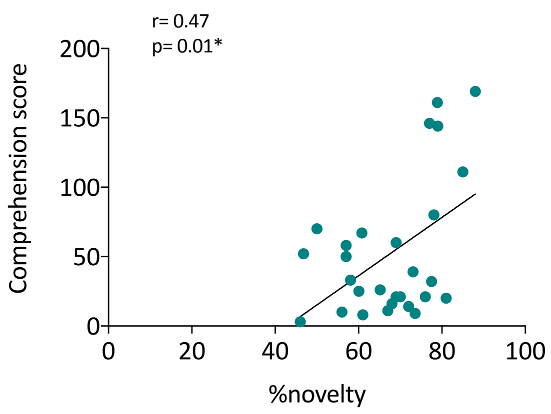
2.5.5.1 Tone cohort

At 12 and 14 months the CDI version Words and Gestures' (IFDC version 1999) was filled by parents of participating infants. The total number of words comprehended as well as the total number of words produced by each infant were calculated.

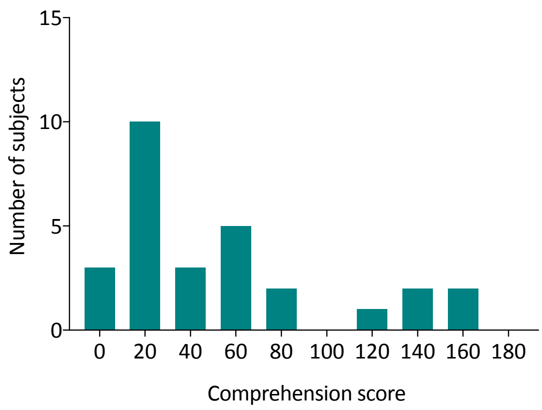
At 12 months (n = 28), vocabulary size ranged between 3 and 169 words comprehended (**Figure 2.19 B**) and 0 and 18 words produced. Pearson's correlation was positive and significant between the comprehension score and the % of novelty effect ($r = 0.47$, $p = 0.01$, cf. **Figure 2.19 A**). No correlation with the production score was found, most likely due to a floor effect, as the production scores were very low at this age.

The positive correlation shows that infants who were better at encoding, retraining in memory and representing information in the visual task also had larger vocabularies. This evidence is line with previous data linking the ability to process novel information with vocabulary (e.g. Thompson et al., 1991; Rose et al., 2009).

Figure 2.19 A. *Correlation between the % of novelty effect and the comprehension score at 12 months (n=28)* **B.** *Frequency distribution of the comprehension score measured across participants at 12 months*



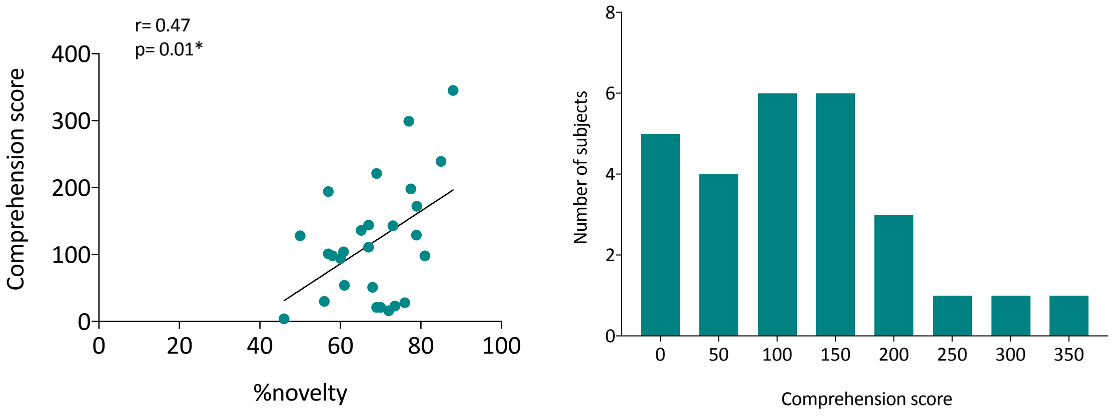
A



B

At 14 months (n = 27), vocabulary size ranged between 4 and 345 words comprehended (**Figure 2.20B**) and between 0 and 81 words produced. The correlation between the % novelty effect and the comprehension score was significantly positive ($r = 0.47$, $p = 0.01$, **Figure 2.20A**). No correlation with the production score was found due to a possible floor effect.

Figure 2.20. A. *Correlation between the % of novelty effect and the comprehension score at 14 months (n=27).* **B.** *Frequency distribution of the comprehension score measured across participants at 14 months*

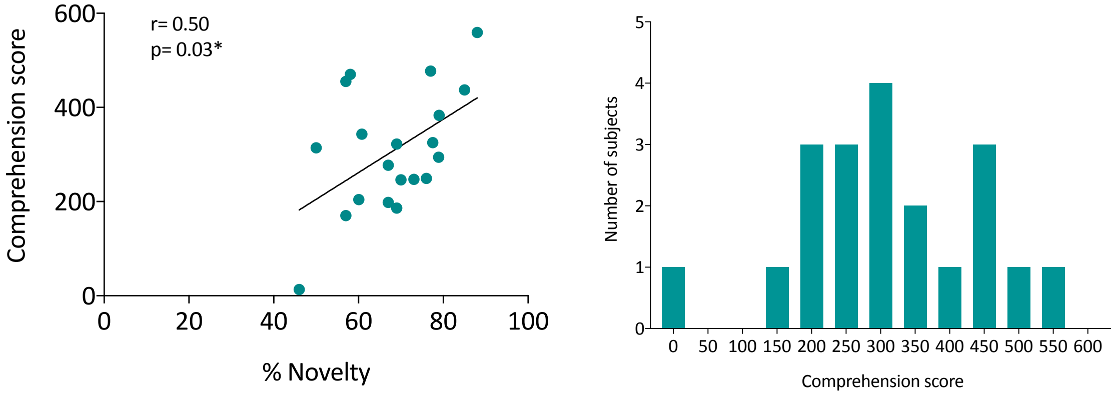


A **B**

At 18 and 24 months, parents filled the Hopkins ‘CDI: Words and Sentences’ and the Hopkins ‘CDI phrases’.

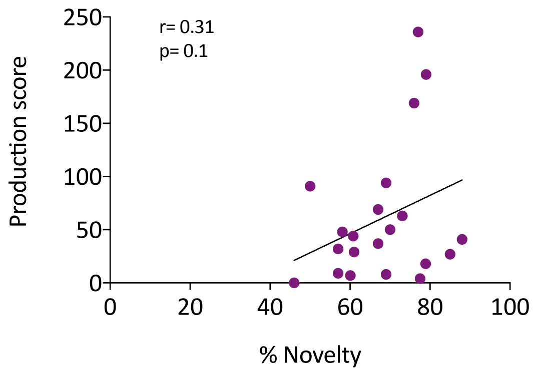
At 18 months (n = 21), vocabulary size ranged between 13 and 559 words comprehended and 0 and 236 words produced. Again, the correlation between the comprehension score and the % novelty was positive and significant (r = 0.50, p = 0.03, **Figure 2.21 A**). Moreover, a positive tendency was also found between the production score and the % novelty effect (r = 0.31, p = 0.1, **Figure 2.22 A**).

Figure 21. A. *Correlation between the % of novelty effect and the comprehension score at 18 months (n =21)* **B.** *Frequency distribution of the comprehension score measured across participants at 18 months*

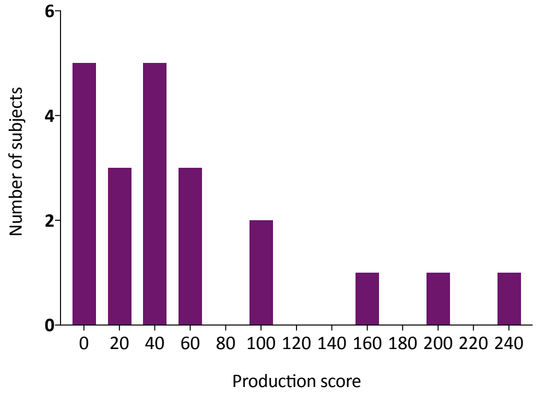


A **B**

Figure 2.22. A. *Correlation between the % of novelty effect and the production score at 18 months (n= 21).* **B.** *Frequency distribution of the production score measured across participants at 18 months*



A



B

At 24 months (n = 12), vocabulary size ranged between 144 and 698 words comprehended and between 9 and 610 words produced. A positive, although non-significant correlation was found between the % of novelty effect and the comprehension score (r = 0.42, p = 0.1, cf. **Figure 2.23**) as well as between the % of novelty effect and the production score (r = 0.28, p = 0.3). Likely, the small sample size (n=12) is responsible for the lack of significance.

Figure 2.23. *Correlation between the % of novelty effect and the comprehension score at 24 months (n=12).*

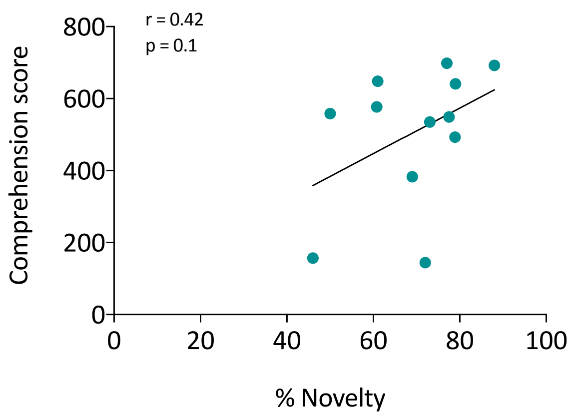
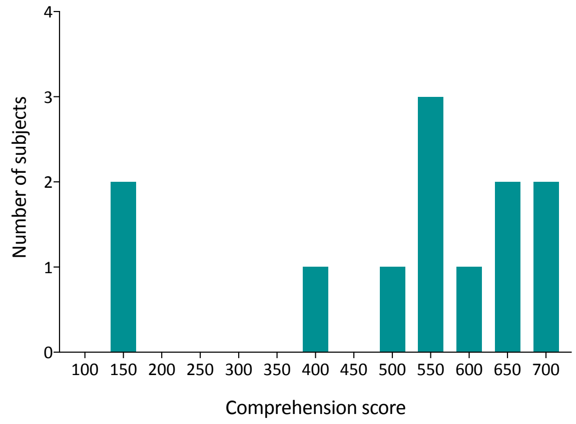
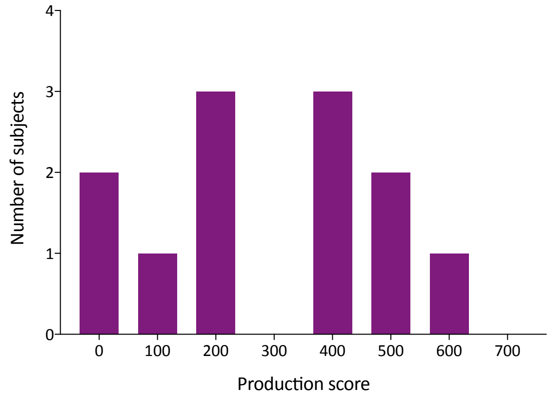


Figure 2.24. A. *Frequency distribution of the comprehension score measured across participants at 24 months (n = 12)* **B.** *Frequency distribution of the production score measured across participants at 24 months*



A



B

Importantly, the analyses have not only revealed the role of novelty effect in predicting language outcomes but they also showed, as reported in the **Tables 2.14** and **2.15**, that measures of vocabulary both in comprehension and production were strongly correlated. As displayed in **Figure 2.25**, only one infant presented a decreased number of words comprehended between 12 and 14 months. Parental vocabulary assessments were thus quite reliable and coherent in this group.

Table 2.14. *Correlation analyses between the % novelty effect and the scores in comprehension*

Comprehension	12 months	14 months	18 months	24 months
Novelty %	.478*	.476*	.479*	.423
12 months	--	.770**	.669*	601*
14 months	.770**	--	.789*	749*
18 months	.669*	.789*	--	954**

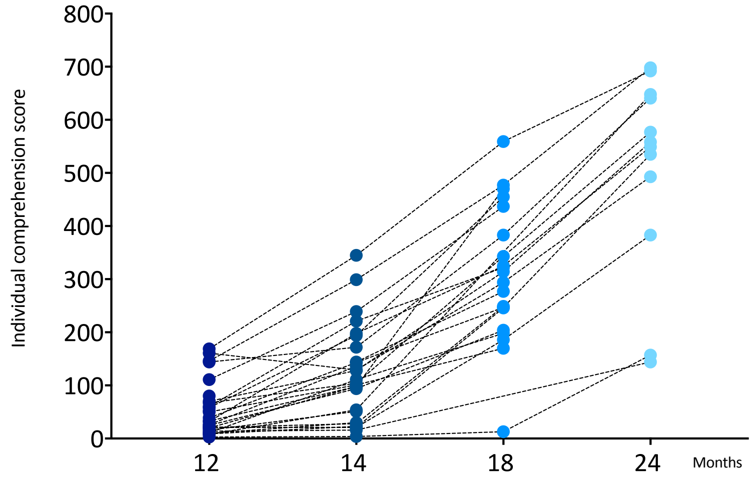
Table 2.15. *Correlation analyses between the % novelty effect and the scores in production*

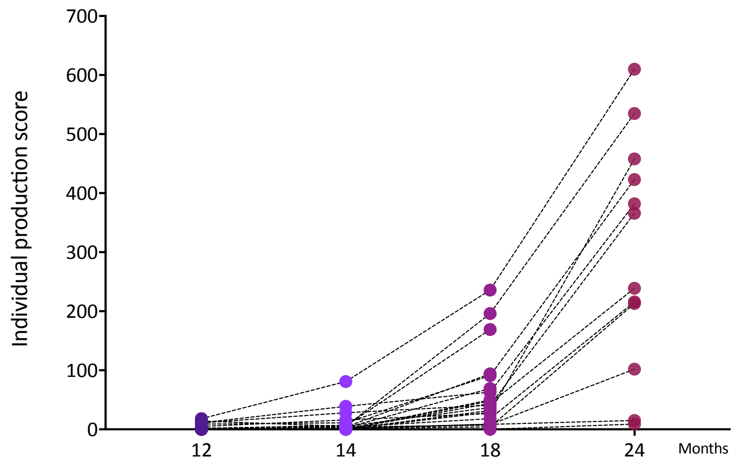
Production	12 months	14 months	18 months	24 months
Novelty %	.323	.355	.31	.279
12 months	--	.810**	.668*	.759**
14 months	.810**	--	.577**	.604*
18 months	.668**	.577**	--	.822**

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 2.25. *Individual vocabulary growth in comprehension (top) and production (bottom) from 12 to 24 months*





To further test the relevance of the % novelty variable as a predictor of vocabulary development, infants in the lowest 5% of the 12-month comprehension score ($M_{wordscompr.}$ = 8.8, SD = 3.96) were compared to infants in the highest 5% of the 12-month comprehension score ($M_{wordscompr}$ = 146.2, SD = 22.2) on their % novelty using an unpaired samples t-test with equal variance not assumed. The difference between the two groups was significant ($t(4.25) = 13.59$, $p < 0.001$) suggesting that infants with the highest and lowest vocabulary scores do indeed differ in their ability to detect novelty. Crucially, infants in the lowest 5% of the 12-month comprehension score also exhibited a different score in the % of novelty effect measured at 9 months: *Low* (M = 61.7, SD = 11.48; *High:* M = 81.58, SD = 4.68, $t(8) = 3.58$, $p = 0.007$).

Interestingly, the infants who were at the lowest and highest 5% of the vocabulary score at 12 months were also in these respective groups at 14 months (*Low*: M = 25.4, SD = 18.6; *High*: M = 236.8, SD = 88.6). The t-test between the two groups' % novelty scores were also significant at this age ($t(4.35) = 5.22$, $p = 0.005$). No other variables reached the level of significance. Due to the smaller number of participants included at 18 and 24 months, this comparison was performed only between 12 and 14 months.

2.5.5.2 Syllable cohort

Identical analyses were performed for the syllable cohort at 12, 14 and 18 months. However, results in this group did not mirror those of the tone cohort. No significant correlation was found between any of the variables and the comprehension or production scores.
Relevantly, when vocabulary scores (both in comprehension and production) were analysed over time, results did not show the same stability. Few of the vocabulary scores were correlated across ages (as presented in **Table 2.16** and **2.17**). This suggests that parental evaluations may not be reliable in this group. Even more revealing is that fact that some parents reported decreasing vocabulary scores between 12 and 14 months, and to a lesser extent between 14 and 18 months (**Figure 2.26**).

Thus, that the lack of correlation between language outcomes and the % novelty effect was most likely due to the fact that the vocabulary measures were not fully reliable and do not thus indicate that % novelty effect is not predictive of vocabulary growth.

Table 2.16. *Correlation analyses between the % novelty effect and the score in comprehension*

Comprehension	12 months	14 months	18 months
Novelty %	- 0.323	.194	.229
12 months	--	.658**	.309
14 months	.658**	--	.792**

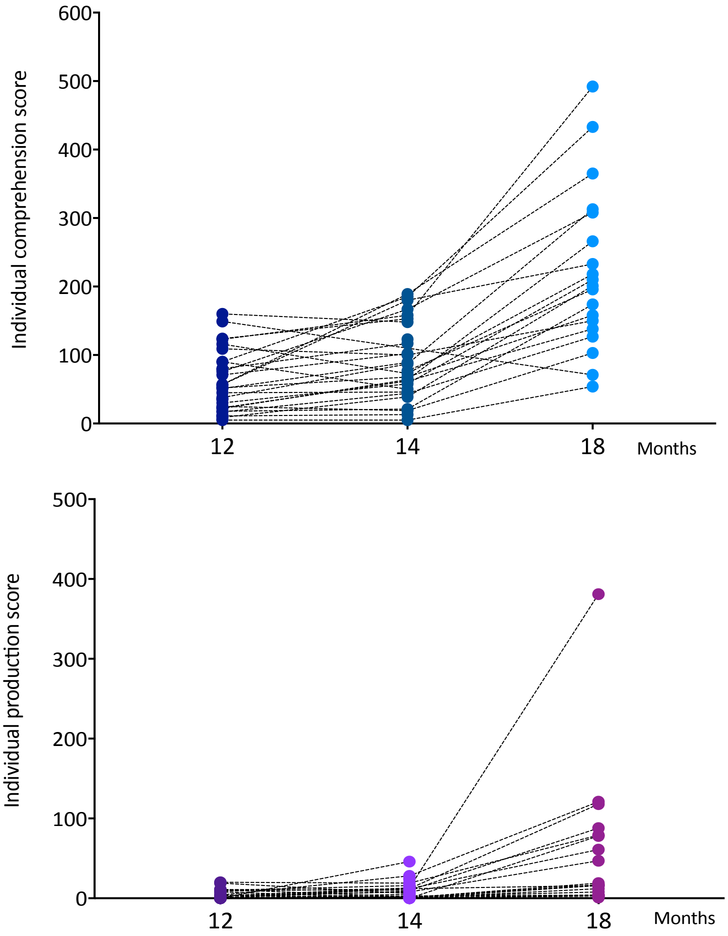
** Correlation is significant at the 0.01 level (2-tailed)

Table 2.17. *Correlation analyses between the % Novelty effect and the score in production*

Production	12 months	14 months	18 months
Novelty %	-0.045	.261	.189
12 months	--	.123	.068
14 months	.123	--	.509*

* Correlation is significant at the 0.05 level (2-tailed)

Figure 2.26. *Individual vocabulary growth in comprehension (top) and production (bottom) from 12 to 18 months*



For a better comparison with the tone cohort, results for the % of novelty effect are nevertheless presented below.

At 12 months (n = 28), a negative tendency between the comprehension score and the novelty effect was found (r = -0.32, p = 0.07, Ns). This result is not only in contrast with those found in the tone cohort, but it is also not in line with the previous literature (e.g. Colombo et al., 1989; Thompson et al., 1991; Rose et al., 2009)

At 14 and 18 months, a non-significant positive correlation was found when the novelty effect was compared with the comprehension (r = 0.19 and r = 0.23) and the production (r = 0.26; r = 0.19) scores.

Figure 2.27. *Frequency distribution of the comprehension score measured across participants at 12 months (n = 28). Values ranged between 5 and 160 words comprehended and 0 and 20 words produced (not reported)*

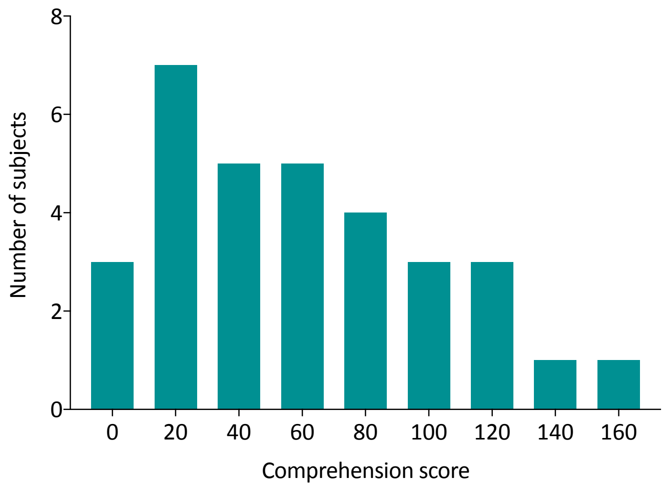


Figure 2.28. *Frequency distribution of the comprehension scores measured across participants at 14 months (n = 27). Values ranged between 5 and 189 words comprehended and 0-46 words produced (not reported)*

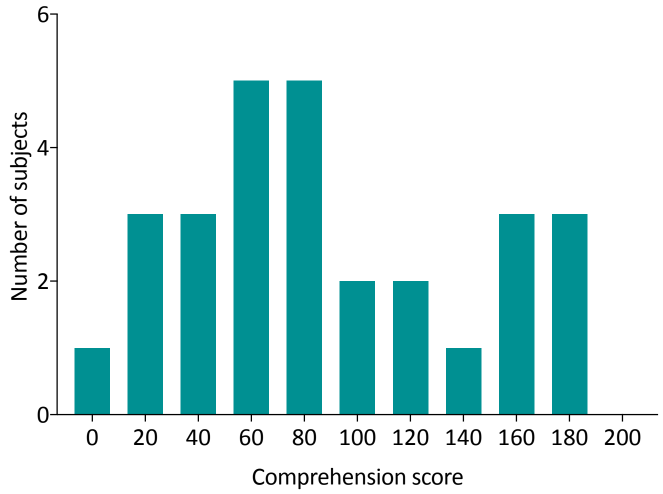
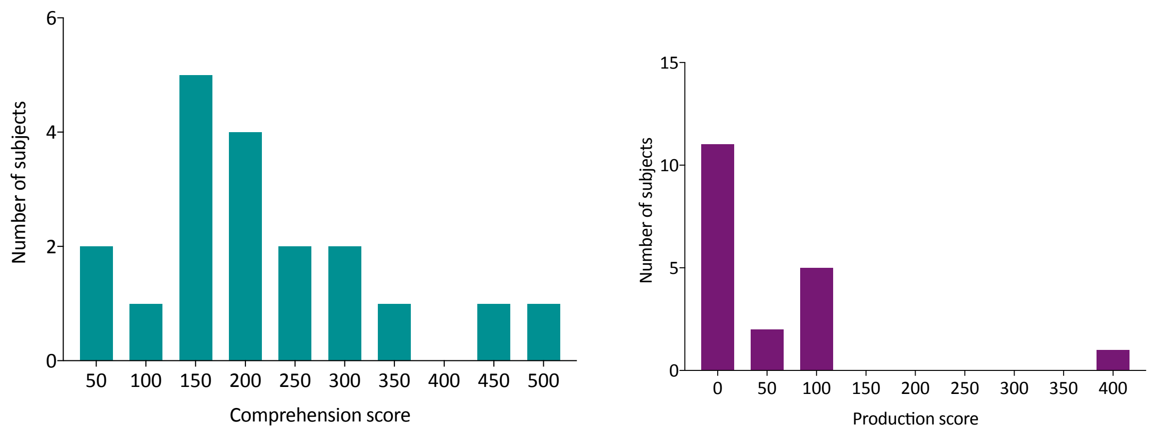


Figure 2.29 A. *Frequency distribution of the comprehension score measured across participants at 18 months (n = 19); values ranged between 54 and 492 words comprehended)* **B.** *Frequency distribution of the production score measured across participants at 18 months (values ranged between 1 and 381 words produced)*



A **B**

2.5.5.3 Standardized questionnaires of vocabulary development: combined results from tone and syllable cohorts

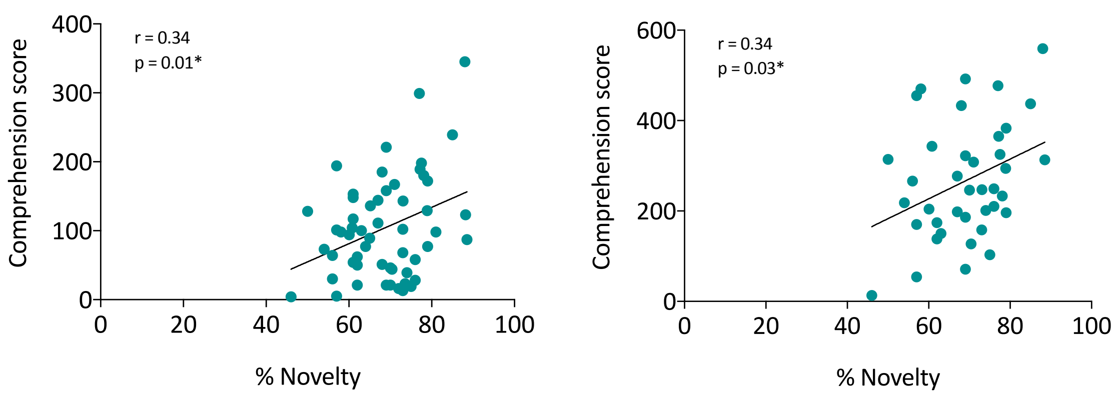
To further compare the reported patterns of vocabulary growth between the two cohorts, in **Table 2.18**, means and SDs of the comprehension and production scores per ages are presented. No differences were found in the production scores, nevertheless the mean number of words comprehended at 18 months was significantly different between the two cohorts: *tones* (M = 308.4, SD = 129.6); *syllables* (M = 221.5, SD = 117.7; t (36.9) = 2.192, p = 0.03).

Table 2.18. *Means and Standard Deviations of the comprehension and production scores measured at 12, 14 and 18 moths between the two cohorts*

	12 months		**14 months**		**18 months**	
Raw score	**Tones**	**Syllables**	**Tones**	**Syllables**	**Tones**	**Syllables**
	(n = 28)	**(n = 32)**	**(n = 27)**	**(n = 28)**	**(n = 21)**	**(n = 19)**
Comprehension	57.7 (49.5)	58.4 (42.6)	118.5 (88.5)	89.7 (55)	308.4 (129.6)	221.5 (117.7)*
Production	3.03 (4.86)	3.4 95.4)	8.7 (16.9)	8.6 (10.6)	60.5 (66.2)	56.6 (88.2)

Importantly, when data from the two groups were pooled together, positive correlations were observed between the comprehension score and the % of novelty effect at 12 months (n = 60): r = 0.13, p = 0.3, Ns, at 14 months (n = 55): r = 0.34 p = 0.01 (**Figure 2.30A**) and at 18 months (n = 40): r = 0.34 p = 0.01 (**Figure 2.30B**). Moreover, at 14 and 18 months a significant correlation between the % of novelty effect the production score: r = 0.31, p = 0.02 (14m) and r = 0.24, p = 0.1, Ns (18m) was found.

Figure 2.30. A *Correlation between the % of novelty effect and the comprehension score at 14 months (n=55).* **B** *Correlation between the % of novelty effect and the comprehension score at 18 months (n=40)*



A **B**

In line with the results obtained with the tone cohort, when the two groups of infants were analysed together, measures of cognitive abilities alone were predictive of linguistic outcomes with stability across time.

2.5.5.4 Standardized questionnaires of vocabulary development: results from typically developing infants and infants at-risk

Lastly, comparisons of language outcomes between typical and atypical participants were performed. Means and SDs of the comprehension and production scores were computed and presented in **Table 2.19**.

Table 2.19. *Means and Standard Deviations of the comprehension and production scores measured at 12, 14 and 18 months between typically developing and at-risk infants*

	12 months		**14 months**		**18 months**	
Row score	**Typical (n = 49)**	**At risk (n = 11)**	**Typical (n = 44)**	**At risk (n = 11)**	**Typical (n = 33)**	**At risk (n = 7)**
Comprehension	56.8 (49.1)	50.81 (29.8)	90.90 (76.7)	120 (63.6)	273.2 (129.3)	227 (139.9)
Production	3.28 (5.30)	3.18 (4.68)	10.04 (15.2)	3.1 (4.08)*	62.0 (81.4)	39 (38.5)

A significant difference was found between the mean number of words produced at 14 months (unpaired t-test): *typical* (n = 44) M = 10.4, SD = 15.2; *at-risk* (n = 11): M = 3.1, SD = 4.08; t (52.57) = 2.63, p = 0.01. Moreover, when 11 infants from the typical group were randomly selected, this difference was still significant: *typical* (n = 11) M = 13.6, SD = 13.02; *At-risk* (n = 11): M = 3.1, SD = 4.08; t (11.96) = 2.54, p = 0.02. Notably, even if not significant, the score in production at 18 months between the two groups showed a trend. This trend remained constant also when 7 infants from the typical group were selected and compared (unpaired t-test): M = 114.8, SD = 85.9, t (8.32) = 2.13, p = 0.06, Ns. Possibly, this comparison did not reach the level of significance due to the small number of subjects present in the groups. Infants at risk thus exhibited a delay in the production abilities.

Moreover, as previously, the % of novelty effect was predictive of language outcomes. A positive correlation was found at 18 months (comprehension: r = 0.84, p = 0.03 and production: r = 0.83, p = 0.01).

2.5.5.5 Discussion: language outcomes

Evaluation of the linguistic outcomes demonstrated that psychoacoustic thresholds at 9 months were not predictive of subsequent linguistic outcomes measured through 24 months of age. Nevertheless, this result could be mostly explained by the limitations encountered during the psychoacoustic evaluation. Contrarily to previous evidence reporting the predictability of early psychoacoustic measures (Benasich and Tallal, 2002), none of the variables here presented correlations with language outcomes.

However, contrarily to the psychoacoustic result, the visual novelty detection task was found to be highly predictive of later vocabulary. Specifically, in the tone cohort, we reported evidence about the link between the response to novelty (in %) and the linguistic abilities over time. These results are in line with previous findings showing the predictiveness of the same variable (e.g. Thompson et al, 1991; Rose et al., 2009). Using a large battery of tasks to measure infants' attention, memory, speed of processing and representational competence, Rose and colleagues (2009) observed that several of these measures could be used as predictors of later language outcomes. For instance, infants' memory and representational competence were related to language at both 12 and 36 months in a concurrent and predictive way.

Indeed, in the tone cohort, the % novelty effect was significantly correlated with the comprehension score at 12, 14 and 18 months. Moreover, a positive tendency was found with the comprehension score 24 months as well as with the scores in production. Relevantly, in this group, measures of vocabulary showed a strong pattern of correlations between them, demonstrating the high reliability of the parental assessments collected over time.

The same results were not replicated with the syllable cohort. No significant correlations were obtained. Since measures of the CDIs were poorly correlated across ages, the reliability of the standardized measures was not as strong as in the tone cohort. Moreover, several fluctuations in individual vocabulary growth were reported by parents between 12 and 14 months

This perhaps less precise parental evaluation might be due to different factors. On the one hand, measuring the comprehension score at 12 months may be more prone to parents' subjective interpretations than scores at later ages. Secondly, we asked families a posteriori whether it was always the same caregiver who filling out the CDIs over time. Only 4 families, to date, reported changes in the caregiver responding to the parental questionnaire between 12 and 14 months. Interestingly, productions scores, often considered a more objective measure, did not present stability across time either. In the syllable group, language outcomes have been evaluated less accurately than in the tone cohort.

Therefore, the non-predictability of the novelty effect within this cohort cannot be interpreted as evidence against the hypothesis that processing efficiency might predict language outcome.

Even if previous evidence already demonstrated the predictiveness of this variable, to our knowledge, no previous data in the literature has shown such a strong and stable relationship between novelty preference and later vocabulary in the same cohort of infants across time.

Relevantly, we chose to use raw scores in comprehension and production for each infant. This choice might represent a more sensitive approach to investigating individual differences and might have led to the stronger and more stable results across time, because applying the standardizing procedure would have caused a significant loss of individual data.

When typical and atypical infants were compared in their language outcomes, no significant differences were found for comprehension scores across time. However, the production score was significantly different at 14 months and 18 months. Measures of production ability could thus be used as an early marker of risk for language delay/impairments.

In a similar vein, a study by Oller and colleagues (1999) reported that the late onset of the canonical bubbling might be predictive of later disorders. In this study, productive abilities of 3400 infants at risk were measured. However, this group also included infants who had been previously diagnosed with significant medical problems that might have accounted for the delay. Results showed that infants who were delayed in the onset of canonical babbling had smaller production vocabularies at 18, 24 and 30 months. This data is relevant in order to consider the stability of production abilities over time when a population at-risk is evaluated.

Even more importantly, Gerrits & de Bree (2009) compared speech perception and production abilities in 3 year-old Dutch children at risk of developmental dyslexia (DD) with a group of infants at risk for SLI and typically developing peers. The authors found that the two groups at-risk (DD and SLI) had poorer production scores than controls. Moreover, the performance of the DD and SLI groups were similar. Our results showed a similar pattern. While not all infants in the at-risk groups were poorer in perception/productive abilities, at the group level, their performance was overall poorer. Gerrits and de Bree argued that DD and

SLI might tap onto similar processing difficulties at early stages of development. Both language impairments may be interpreted within a multi-risk model where cognitive processes and genetic factors have to be strongly considered.

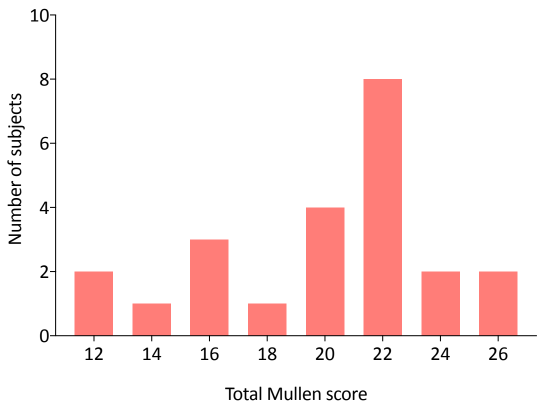
2.5.6 Mullen scale of Early Learning

Lastly, results of the Mullen Scale (Mullen, 1995) are presented in this section. The test was administered to 23 infants (10 from the tone cohort and 13 from the syllable cohort). Not all families had the possibility to bring their toddlers back to the laboratory, resulting in these small sample sizes. Infants were tested between 18 and 20 months (Mean Age 20m, 2 d).

No effect of gender was found between females ($n = 10$) and males ($n = 13$) performing the Mullen Scale (females: $M = 19{,}8$ $SD = 4.13$; males: $M = 19.6$, $SD = 4.11$). Importantly, the age effect was marginally significant ($r = 0.27$, $p = 0.057$), indicating that older infants tended to perform better during the test.

The histogram of the Mullen scale total score, calculated by summing the raw scores of each single domain (visual/ fine and gross motor/ receptive and expressive skills) is presented in **Figure 2.31** (score between 12 and 26 points). At the group level no significant variability in the total score was present. Perhaps, at these early stages of development, cognitive tests highlight significant differences only when infants are considerably delayed with respect to peers.

Figure 2.31. *Frequency distribution of the Mullen total score measured across participants (n = 23)*

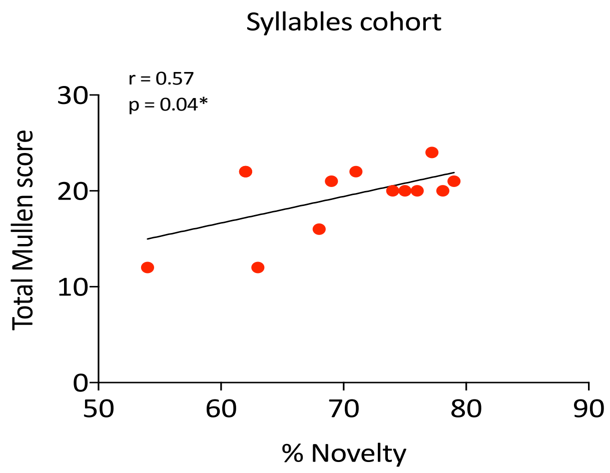


We calculated correlations between the Mullen score and all the variables measured in the two behavioural tasks. This particular analysis was important to evaluate whether early behavioural measures could be potentially used as early predictors for later general development (cognitive and linguistic).

The novelty effect showed a significant correlation with the Mullen score for the syllable cohort (n = 13; r = 0.57, p = 0.04, cf. **Figure 2.32**), but not for the tone cohort (n = 10; r = 0.31).

Furthermore, no links between the Mullen receptive vocabulary score and the CDI score in comprehension were found (r = 0.05), nor in the Mullen expressive vocabulary score and the CDI score in production (r = -0.10).

Figure 2.32. *Correlation between the % of novelty effect and the Mullen total score (syllable cohort n = 13)*



2.6 Discussion

In this longitudinal study perceptual abilities were investigated in two groups of 9-month-old infants followed longitudinally. Specifically, a RAP paradigm was run to evaluate whether the auditory temporal threshold measured at early stage was predictive of later language measures. Moreover, the language-specificity of the processing was investigated using linguistic and non-linguistic sounds in two separate cohorts. In addition, a visual novelty detection task was performed as a control for general cognitive skills. Subsequently, infants' vocabulary was assessed at 12-14-18 and 24 months and a cognitive test (Mullen scale) was performed at 18-20 months as another control measure for early processing competence.

Results have demonstrated that the auditory temporal threshold was not a sensitive predictor of subsequent language development. This was in contrast with previous evidence showing that differences in basic auditory processing abilities are linked to subsequent receptive and expressive vocabulary (Benasich & Tallal, 2002). However, we discussed methodological limitations that may have led to this result. Moreover, considering the high variability in the literature regarding infants' auditory threshold (see, for instance, Benasich and Tallal, 1996; Trehub et al., 1995 and Bishop & Sutcliffe, 2005 for a comparison), this study emphasizes the necessity of implementing the procedures to measure auditory threshold discrimination in infancy.

Participants in the two cohorts were split into three sub-groups based on the individual performance ability and variation on auditory discrimination during the task. Three profiles were established: the *good*, the *poor* and the *inconsistent/unstable* performers (based on Moore et al., 2008). This division highlights the fact that individual variability arose from different auditory discrimination efficiency together with fluctuations in general cognitive processes (e.g. attention) during the task. In fact, changes in attention during the performance were more frequent in participants that exhibited a higher auditory threshold. In line with previous evidence (e.g. Moore et al., 2008) this might suggest that infants who are better processors of acoustic information are also those who show more stability in the general attentional level.

Contrarily to the initial expectation, when the acoustic performance was compared between linguistic and non-linguistic sounds, no difference was found. However, performing syllable discrimination may still have been more challenging, as suggested by two measures. First the rejection rate was higher for the syllables (42.8%) than for the tones (30.4%), meaning that a bigger proportion of infants was not able to reach the test phase. Moreover, comparing the two groups of *inconsistent/unstable profile*, they did not behave in a similar manner. In the syllable cohort, this profile not only exhibited several fluctuations in the attention during the task, but infants had significantly higher thresholds than infants with inconsistent profiles in the tone cohort.

In addition, possibly due to the methodological limitations encountered in the psychoacoustic evaluation, no direct relations between the psychoacoustic performance and the visual novelty detection task could be observed. This finding was in contrast with previous results where authors hypothesized that “the infant who can integrate and process information efficiently across domains is likely to acquire a larger, more detailed knowledge base over a particular time period than an infant who processes more slowly” (from Benasich and Tallal, 1996, p. 352).

Furthermore, we also explored whether measures of cognitive abilities alone were predictive of linguistic outcomes with certain stability across time. Specifically, measures of habituation can be interpreted as how fast and well infants encode stimuli, create a representation of them and retain them in memory (Rose, Feldman & Wallace, 1988).

Crucially, in the present study, we reported evidence about the direct link between the response to novelty and the linguistic abilities over time. This result is in line with previous findings showing the predictiveness of this variable (e.g. Thompson et al, 1991; Rose et al., 2009). Nevertheless, to our knowledge, no previous study has shown such a strong stability of this measure in the same cohort of infants across time. In the tone cohort, the % of novelty effect was predictive of the receptive and expressive vocabulary between 12 and 24 months. It needs to be noted, though, that the same results were not found in the syllable cohort, even if the effect was present when the two cohorts were analysed together. Likely, imprecision in caregivers’ evaluations of vocabulary led to this weaker effect.

One of controversial issue is whether cognition is stable throughout early development. Some authors in the past proposed that there is a certain degree of discontinuity in early cognition since cognitive measures at one stage were found to be unrelated to individual ability at later stages (Kopp & McCall, 1982; Bayley, 1949; McCall; 1979). In addition, studies have proposed that there is a huge variability in cognitive measures difficult to interpret or to link directly with later outcomes (Bayley, 1969).

At the same time, other studies tried to analyse the stability of these measures during development. For instance, using a battery of several tasks, it has been shown that cognitive abilities influenced each other and the later cognitive development (Rose, Feldman, Jankowski et al., 2005; Rose et al., 2007).

The present results are not conclusive since the results of the Mullen Scale revealed that the % of novelty effect, measured at 9 months, was correlated with the total score achieved in the cognitive test between 18 and 20 months only for the syllable cohort but not for the tones. Moreover, at the group-level the effect was not significant. Hence, perhaps, the cognitive test that it was administered here did not strongly highlight individual variability between performers.

Finally, one of the main purposes of this work was to investigate differences in auditory, cognitive and linguistic behaviour between typical and atypical infants. Importantly, we identified three differences, which could provide potential behavioural markers of risk for language problems.

First, infants at-risk showed slower discrimination abilities and higher fluctuations in the attention level (e.g. presence of inattentive trials). They learned the operant conditioning procedure slower and less efficiently than their typical peers, showing slower processing efficiency in both the auditory and the visual modalities.

Moreover, the magnitude of the novelty effect was significantly different in the two groups, meaning that atypical infants were less efficient in encoding, retaining and processing the information. This finding is in line with the previous literature suggesting that the % of novelty effect, measured in similar tasks, can be use as a reliable predictor of later cognitive outcomes both for typical and atypical populations. By examining infants' abilities during visual recognition memory performance at 7 and 12 months, Rose and Feldman (1995) found that these measures were stably predictive of specific cognitive outcomes at 11 years in both pre-terms and full terms. In addition, a meta-analytic review performed by McCall and Carriger (1993) revealed that infants' habituation and recognition memory outputs, measured during the first year of life, were sensitive predictors of later IQ at 8 years of age. Moreover, these measures resulted also predictive of specific differences between typical and atypical populations.

Our findings seem overall in line with the existing literature showing that infants at-risk of language deficits/delays might have greater difficulties in performing both the auditory processing and the visual novelty detection task. Even if the not all the results matched previous evidence, informative differences were found by comparing the two samples.

Lastly, when compared to typical participants, atypical infants showed a lower score in the expressive vocabulary at 14 and 18 (Ns.) months. Hence, we also advance the hypothesis that production ability could represent a potential early diagnostic tool for language delays.

In sum, this study contributes to a better understanding of whether auditory and cognitive measures evaluated at early stages can be used as predictors of later language outcomes. Our findings imply that, by combining linguistic and cognitive measures, a more subtle understanding of individual language development may be obtained. We have also identified possible measures that might be useful in early identification of infants at risk of language delay/impairments.

Chapter 3:

Bootstrapping word order and lexical categories: the role of word frequency

3.1 Introduction

Infants are faced with the task of learning the sounds of their native language, connecting specific sound sequences to meaning, finding phrases and words in the continuous speech stream and discovering the relationship between these units. The previous chapter discussed how basic auditory and visual perception may contribute to language learning. The current study, by contrast, investigates a mechanism that may help infants tune into grammar.

In the General Introduction, bootstrapping mechanisms have been extensively discussed as mechanisms that guide infants from surface cues of language to lexical and grammatical knowledge. Word frequency has been evoked as one possible cue to categorize functors and content words. The *frequency-based bootstrapping hypothesis* (Gervain et al., 2008; Bernard & Gervain, 2012) relies on the language universal division of labor between these two lexical classes (Chomsky, 1995; Fukui 1986; Abney, 1987). The two classes differ in their frequency of occurrence, phonological make up and lexical function and occupy different positions in syntactic units.

Language typology traditionally distinguishes between two types of languages in terms of basic word order. In Verb-Object (VO) or functors-initial languages, functors precede the content words. Examples of VO languages are French, Italian and English (e.g. English: *in London*, *he runs*, French: *à Paris* (in Paris), *il court* (he runs). By contrast, in OV or functors-final languages, like Turkish, Basque and Japanese (e.g. Japanese: *Tokyo ni* (Tokyo to), Basque: *garren atzean* (flame behind), function words are placed in final position (Nespor & Vogel, 1986). The ability to identify functors and content words can help infants establish the basic word order type of the native language.

Table 3.1. *Summary table of the main differences between function and content words*

	Function words	Content words
Function	signalling grammatical relations	carrying lexical meaning
Phonological properties	reduced, minimal	carry prosodic prominence
Distributional properties	high frequency	mid-to-low frequency
	fixed position in phrases	
Lexical properties	Closed-class	Open-classes
	(No new items)	(New items can be added)

In this chapter, we report five artificial grammar learning experiments asking whether preverbal infants are able to use these word frequency as a establish the lexical categories of functors and content words and learn their relative order in the native language.

Content words (e.g. nouns: *giraffe, table, dress, etc.,* verbs: *drink, listen, play, etc.,* adjectives: *powerful, amazing, good, etc.)* come in large, open classes: new items are added to the lexicon every day (e.g. *iPad*, *Brexit, USB,* etc.), whereas functors (e.g. articles: *the, a, etc.*; personal pronouns: *I, you, he, she, etc.*; prepositions: *of, on, up, etc.*) constitute closed classes, into which no new items can be added without a major language change. This is insightfully illustrated by Lewis Carroll's Jabberwocky poem: " *Twas brillig, and the slithy toves / Did gyre and gimble in the wabe* […]". In this citation the grammatical structure is maintained by function words, whereas content words are replaced by novel tokens. In addition to their distinctive functional properties, the two classes also differ in their frequency distribution and phonological properties.

Function words are phonologically reduced or minimal. They tend to have fewer syllables and simpler syllable structure with respects to content words as well as weaker prosodic prominence (shorter, less prominent in pitch and weaker in intensity) (Nespor & Vogel, 1986; Selkirk, 1984). Different corpus studies have provided evidence for the existence of these acoustic differences. By comparing English and Chinese infant-directed speech, Morgan, Shi, and Allopenna (1996) showed that functors differ in a series of specific phonological measures (e.g. number of syllables, syllable complexity, vowel duration, amplitude etc.) that make them more reduced than content words, although the measures in which the two classes differed most varied between the languages.

Infants seem to be sensitive to these acoustic/phonological differences. Indeed, already newborns detect the different acoustic saliency of the two word classes (Shi, Werker, & Morgan, 1999). Furthermore, 6-month-old infants are able to discriminate content and function words across two languages that are radically different (Chinese and English). Chinese-learning infants prefer to listen to content words over functors in English. This indicates that language-general acoustic and phonological cues are sufficient for an initial discrimination of the two classes even without prior exposure to the language (Shi & Werker, 2001, 2003). Taken together these findings suggest that universal mechanisms are present since birth in order to distinguish the acoustic properties of the two categories and these properties might help bootstrap a rudimentary representation of them.

The differential frequency distributions of the two categories are also language-universal. Functors have a much higher token frequency than content words. This was confirmed in several corpus studies. Cutler and Carter (1987) and Cutler (1993) reported that functors represent the 59% of the word tokens of their corpus, while they constitute only about 1% of all the word types. Moreover, a study by Kucera & Francis (1976) observed that there is little overlap in the frequency distributions of functors and content items in English since the 50 most frequent lexical items in English were found to be function words. Japanese and Italian cross-linguistic analyses of infant-directed speech showed similar results. In fact, in the study of Gervain et al. (2008) a corpus analysis was performed in order to evaluate the frequency of occurrence of function words and content words in infant-directed speech across two languages (Italian and Japanese). As expected, individual functors occurred more frequently than individual content words in both languages. Specifically, calculating the frequency distribution of the 100 most frequent words, in Italian 83.13% (50.93% of the whole corpus) and in Japanese 79.73% (37.83% of the whole corpus) were functors. This finding is in line with the frequency distribution of words expressed by Zipf's law (Zipf, 1953): in languages there exist a few very high frequency words, typically functors, and a large number of much less frequent words. The authors also looked at how often frequent and infrequent words appear in initial and final positions of utterance boundaries in the two languages to test whether the position of frequent words offers a perceptually available correlate of basic word order type. The results confirmed the predictions. Italian had significantly more utterances starting with a frequent word, whereas Japanese had a higher number of frequent words in utterance-final position. This analysis confirmed that the input contains statistical information about word order, which pre-lexical infants might be able to extract.

Recently, non-human animals' ability to track the frequency of words in speech was investigated. Rats were trained to press a lever to obtain pellets in response to different sequences that either systematically alternated between frequent and infrequent elements, similarly to the stream used in Gervain et al. (2008), or were presented randomly. Animals showed an increasing level of response after the alternation sequences and a decreased response to the random ones, indicating that they were successfully trained to recognize the structured stream. They were then tested, similarly to infants, on frequent-initial and frequent-final chunks from the training stream. Rats systematically selected frequent word initial test items, showing that they were able to differentiate between frequent and infrequent elements (Toro et al., 2016). However, they paid no attention to the structure of the test items, and selected all items starting with a frequent word, even if no alternating structure was present, e.g. if the initial frequent word was followed by another frequent word. The presence of this skill in mammals suggested that frequency-based bootstrapping is a general learning mechanism shared across species. However, attention to structure seems to be specific to humans. This meshes well with the fact that pre-lexical infants use this strategy to infer the word order of the native language.

Overall, the frequency-based discrimination of functors and content words lexical constitutes a potentially powerful initial mechanism for infants to acquire the most basic building blocks of language as well as their relative order.

3.1.1 Bootstrapping the word order: experimental evidence

Several studies in the literature have shown that pre-lexical infants, as well as adult participants, are sensitive to the relative distribution of frequent and infrequent items in the linguistic input. The experimental evidence that is presented in this section is based on a specific paradigm called Artificial Grammar Learning (AGL), dating back to Reber (1967). An Artificial Grammar is a set of rules over (categories of) lexical items. Participants are typically unaware of the rules, so AGL usually taps onto "*implicit learning*" mechanisms. Participants are presented first with a training phase (also called familiarization phase) during which they are exposed to strings from the grammar. They are then tested on different sequences, e.g. grammatical vs. ungrammatical, familiar vs. novel ones etc. If participants are able to extract the regularity from the input during training, then they can discriminate between grammatical and ungrammatical test items.

To investigate the frequency-based bootstrapping hypothesis, Gervain et al. (2008) tested Japanese and Italian 8-month-olds with an AGL task in order to test whether they had different expectations (opposite looking preferences) for word order. An AG was constructed by repeatedly concatenating a four-syllable-long basic unit: AXBY, where A and B represented constant syllables, while X and Y came from two categories containing 9 syllable tokens each. Thus, an alternating sequence of frequent (A & B) and infrequent (X & Y) syllables was obtained, mimicking the statistical occurrence of functors and content words in natural language. The stream was ambiguous in its structure, because the initial and final 15 s of this familiarization stream were ramped in amplitude. This resulted in an ambiguous underlying structure, since the basic unit could be identified either as having a frequent-initial (AXBY) or a frequent-final (XBYA) order. The stream was flat in prosody, thus word frequency was the only available cue. During the test phase eight different items were presented, four with an AXBY structure and four with an XBYA. Japanese infants preferred test items that matched the order of the functor-final language (frequent items in final position, XBYA) whereas Italians the functor-initial one (frequent items in initial position, AXBY).

The validity of the frequency-based bootstrapping mechanism was also established by studies with adult participants. Studies with monolinguals and bilinguals examined whether adults also rely on word frequency when they have to segment a novel input from an artificial language. Gervain (2007) and Gervain et al. (2013) compared adult monolingual speakers of languages with different word orders (French and Italian, VO; Japanese and Basque, OV; Hungarian mixed order) in the adult version (familiarization + 2 alternative-forced choice test) of the same AG as in Gervain et al. (2008). Speakers of OV languages exhibited a significant preference for the test items that were coherent with their native word order (frequent-final items). Speakers of VO languages did not show a statistically significant preference for the frequent-initial order as compared to baseline, but performed significantly differently from OV speakers. Furthermore, using the same paradigm, de la Cruz-Pavía et al. (2014) tested Basque (OV) / Spanish (VO) bilinguals. Bilinguals were in general more flexible, being able to extract both word orders and switch between them, depending on the language in which they were addressed during the experiment.

Word frequency information is correlated with prosodic information at the phrasal level. In VO / frequent-initial languages, e.g. French, phrases exhibit a short-long pattern, since prosodic prominence is realized with an increased duration on the stressed vowel of the prominent element, the infrequent the content word, whereas in OV / frequent-final languages, such as Japanese, phrasal prominence is typically realized with increased pitch and/or intensity on the stressed vowel of the prominent infrequent content word, so phrases tend to have a high-low or strong-weak pattern (Nespor et al., 2008). Some existing studies, therefore, investigated infants' performance when word frequency and prosody were combined. In a study with French monolingual infants, Gervain and Bernard (2012) investigated whether infants were able to process and integrated the two cues simultaneously. Two conditions were presented. In Condition 1, the two cues were aligned in the familiarization stream: non-prominent frequent words were short (240ms) and the prominent infrequent words were long (320ms)], mimicking natural language. In Condition 2, the two cues were in conflict, i.e. frequent words carried lengthening, infrequent ones did not. This pattern is unusual in languages, as frequent words are functors and thus phonologically minimal. Those carrying prominence, e.g. lengthening, are the content, i.e. infrequent words. Nevertheless, this unusual pattern could only be detected if infants processed the two cues simultaneously at the individual lexical level. If the cues were processed separately and/or at a more global level, even Condition 2 could be perceived as well-formed and native-like, as it contained a durational contrast, following French prosody. Results showed that infants preferred the frequent-initial pattern when the cues were aligned at the level of lexical items (Condition 1), but no significant preference was found in the misaligned condition (Condition 2). This seems to suggest that infants expect prosodic cues and word frequency to be aligned at the lexical level.

Another recent study aimed to understand whether bilingual infants might benefit from these cues. Gervain & Werker (2013) argued that when infants are exposed to two languages with opposite word order, word frequency alone is not sufficient as a cue, because both frequent-initial and frequent-final patterns appear in the input, and phrasal prosody might provide an additional cue. In an artificial grammar task similar to the previous ones, they tested English (VO) – Japanese (OV) bilinguals and a group of English monolingual. A first group of bilinguals was tested in condition 1. Infants were familiarized with an OV prosody in which infrequent words were higher in pitch (224 Hz) with respect to frequent ones (200 Hz). In condition 2 a second group of bilinguals was familiarized with a VO prosody in which frequent words were shorter (120 ms) than the infrequent ones (144 ms). In both conditions bilinguals showed the expected preference for the word order that corresponded to the prosody. This suggests that bilingual infants may indeed use prosody as a cue to word order. A group of English monolinguals was familiarized with the same OV condition (condition 3), while another group heard a no-prosody condition (condition 4) in which the stream was flat, similarly to Gervain et al. (2008). English being a VO language, infants were never exposed to the OV prosody, resulting in a null preference. In the no prosody condition, they showed the predicted preference for the frequent-initial items. The results suggest that bilingual infants are highly sensitive to both cues, in line with bilingual adults' behavior (de la Cruz-Pavía et al., 2014). Furthermore, monolinguals are sensitive to the conflict between the non-native OV prosody and word frequency, but prosody is not sufficiently strong to override word frequency and push infants all the way to an OV word order preference.

Taken together, these studies show that infants are able to use both prosodic and frequency information to acquire word order. As 8-month-old infants are able to integrate word frequency and phrasal prosody at the level of lexical categories, this bootstrapping mechanism is pre-verbal and thus, independent of vocabulary learning.

What is the exact nature of the representation of word order that infants extract in these studies? They may simply represent the relative position of frequent and infrequent words. A more interesting possibility is that they actually represent the relative position of functors and content words. However, this has never been tested directly. The purpose of the current experiments is to test whether infants are able to use word frequency for early lexical categorization, i.e. to establish the categories of functors and content words.

3.1.2 Functors make language learnable

The hypothesis that functors are necessary in order to learn the rules of language has been investigated in several artificial grammar studies. Language is assumed to be unlearnable without this word class, as functors sere as anchor points with respect to which the position of other items can be encoded.

Braine (1966) was one of the first to have experimentally explored this hypothesis. He was interest in understanding how constant elements in speech influence language learning. What he assumed to be relevant was not the ordinal position of words but the specific position of a single word with respect to the others. In an experiment, he presented 9-year-old children with a set of sentences containing a constant marker element and various content words. Children had to learn the position of the content words with respect to the functors. Two conditions were tested. The content words were either immediately adjacent to the marker or one position away from the marker. Content words were acquired easily in both positions, as the constant markers helped children extract the other words.

Later, Green (1979) presented to adult participants different sets of sentences in which i) markers occurred in a systematic way, ii) marker were “useless” as they occurred randomly or iii) no markers occurred. As predicted, participants performed significantly better in the condition where markers were presented systematically than in the other two conditions. It is important to note that the study was conducted with adults, who have good knowledge of language, thus pre-acquired capabilities might have played a role.

Similar results were obtained by Morgan et al. (1981, 1987). Authors tested the role of free functors and suffixes separately, using three artificial grammars: (i) no markers (only content words) (ii) inconsistent markers (functors appeared randomly) or (iii) consistent markers (functors indicated phrase boundaries). In both experiments (free functors vs. suffixes) participants demonstrated that the presence of markers represented the variable that facilitates the learning of complex aspects of syntax, e.g. linguistic dependencies. Nevertheless they performed better in the condition where functors or suffixes were consistent.

Following these results, Valian and Coulson (1988) investigated how adults’ performance in learning artificial languages that differs in the number of markers (functors) and content words. Two dialects were created using non-sense function and content words, then markers frequency was manipulated. The dialect 1 contained 2 marker tokens and 12 content tokens, whereas the dialect 2 contained 4 marker tokens and 6 content ones. Two experiments were performed. In experiment 1, the learnability of the two dialects was evaluated, showing that the high frequency dialect was easier to learn respect to the low frequency one. In Experiment 2, where reference fields were added (coloured dot labels and stylized round patterns), both languages were learned but the high frequency exhibited an overall better performance. The authors discussed that in both experiments the high frequency markers condition always facilitated the acquisition of the grammar, as markers were used to build correct constructions.

Overall these findings support the “marker/anchoring hypothesis”, arguing that all languages are provided with a set of words, like functors, that are associated with specific syntactic constructions. Without the presence of these constant elements, it wouldn’t be possible to learn language.

3.1.3 Is the early perception of functors consistent with their late production?

Functors thus seem to be crucial for language learning. Paradoxically, however, infants produce functors late. Is this compatible with the hypothesis that functors are privileged anchor points for processing and learning grammar? The telegraphic nature of children's first multiword productions (Guasti, 2002) led to the idea that language development initially lacks function words. Some authors attributed this to the fact that functors are more abstract and grammatically more complex than content words (Brown, 1973), whereas others proposed that is a biologically predetermined later developmental stage in acquisition (Radford, 1990). It is in fact important to distinguish sensitivity/perception and production. Studies assessing early perceptual abilities suggest that a rudimentary representation of functors is in place well before the ability to produce them. The omission of functors does not reflect a limitation on perception or encoding.

Empirical evidence to support this is abundant. For instance, infants perform better in tasks in which the instructions are not telegraphic but contain functors (Shipley, Smith & Gleitman, 1969). Also, 2-3-year-old English children, when asked to imitate non-sense words, omit shorter, weaker and unstressed words (similar to funtors) more often than longer and stressed non-sense words (similar to content words; Gerken et al. 1990). Moreover, they imitate non-sense content words with more accuracy when those are accompanied by real function words. This confirms the idea that functors help categorize content words serving as anchoring points for grammatical structure (Gerken, Landau, & Remez, 1990).

In a later study, similar results were obtained when sentence comprehension was analysed. 24-month-olds were exposed to imperative sentences in a picture selection task: different pictures were presented and children had to choose the one that matched the sentence they heard. Functors in the sentences could be (i) grammatically correct, (ii) ungrammatical or (iii) totally omitted. Children's performance was significantly better when grammatical functors were used indicating that, even if their production is not yet in place, function words are perceived and necessary in the early stages of comprehension (Gerken & McIntosh, 1993).

Overall these results lead to two conclusions. First, infants produce functors later than content words, but this delay in early production can be attributed to production constraints. Secondly, children have a detailed representation of function words even if they might have problems in producing them.

3.1.4 Functors in early segmentation skills and word learning

Infants are exposed to a continuous speech stream that they have to break into phrases and words. As it was discussed in the previous sections, infants are able to group functors into a distinct category even at birth (Shi, Werker, & Morgan, 1999). Once they know some functors, infants could start to learn the general structure using known functors as anchors for segmenting words and discovering their syntactic properties. In a study by Höhle and Weissenborn (2003) 8-month-old German infants exposed to continuous speech were able to segment functors from the stream. Between 8 and 11 months infants can use high frequency but not low frequency functors to segment adjacent content words (Marquis & Shi, 2012; Shi & Lepage, 2008; Shi, et al. 2006a; 2006b). Neurophysiological evidence suggests that 11 months participants showed different ERP responses while they were listening to continuous speech versus continuous speech in which a tone replaced function words (resulting in a distortion of the acoustic features) (Shafer, Shucard, & Gerken 1998). Furthermore, 14-16-month-olds group together similar functors (e.g. determiners vs. pronouns) and they are able to categorize novel words as nouns only when a familiar determiner is placed in the correct position (Shi & Melançon, 2010).

Function words also affect word learning. In a context in which novel words (e.g. *larp*) are presented, infants are more likely to categorize the word as a noun if the novel word is preceded by functors that support nominalization ("Look! The man is waving **a** *larp*"), whereas if the functors indicate the use of a verb ("Look! The man **is** *larping* a balloon"), the novel word is more likely to be used as referring to a picture of an action (Waxman, Lidz, Braun, & Lavin, 2009). Furthermore, in an artificial grammar/object labelling task, 17-month-old infants familiarized with a foreign language more readily associated infrequent words than frequent ones with a possible referent, suggesting that they expect infrequent words to have semantic content (Hochmann, Endress, & Mehler, 2010).

3.2 Objectives of the present study

In the present study we aim to extend the existing research on the acquisition of functors and content words. Previous accounts have already shown that infants rely on the different distributional and phonological correlates of function and content words to bootstrap the word order of their language, but there is not much evidence regarding whether they actually establish distinct lexical categories for function and content words at an early age. The existing experimental evidence shows that infants are sensitive to frequent elements in their input but it is still unclear whether they identify them as function words.

We tested the above hypothesis by examining whether infants treat frequent words as coming from closed classes and infrequent ones as coming from open classes, a respective characteristic of functors and content words. We tested this with five different experiments. Experiment 3 attempted to establish French learning-infants' preference for the frequent word initial order on the basis of frequency cues, extending findings (Gervain et al. 2008) to French. Experiment 4 tested infants' sensitivity to the open class property of content words, by replacing the infrequent words in the test items with new ones. Analogously, the sensitivity to the closed class property of function words was tested in experiment 5 by replacing frequent words with novel ones in the test items. To make sure that any significant results from the above experiments arise from differences in how infants categorize frequent and infrequent elements, and from infants' inability to remember infrequent words, experiment 6 tested the recognition of infrequent words. Finally, in experiment 7, we tested whether infrequent items alone can be use to encode information about word order.

3.3 Materials and Methods of Experiments 3-7

The Head turn Preference Paradigm (HPP) (Kemler Nelson et al., 1995: Saffran, Newport et al., 1999) was used to test this hypothesis. This experimental method measures infants' looking behaviour to assess preferences for and/or discrimination between at least two different types of auditory stimuli. With this technique infants learn to orient visually to a sound source while they maintain a response (a head-turn) when stimulation is contingent on their behaviour (Kemler Nelson et al. 1995). Infants' attention is assessed evaluating the length of time their head is turned towards the sound source.

3.3.1 Procedure

The experiments were conducted in a quiet testing booth with three screens, one in front of the infant and one on each side, on which visual attention getters (e.g. videos of looming circles imitating blinking lights) were played (one on each side; **Figure 3.1**). Below each side screen, loudspeakers were placed for the presentation of the sound stimuli. Infants were seated on a caregiver's lap, sitting on a chair in the middle of the booth. The caregiver listened to masking music in order to avoid influencing the infant's response. Each experimental session consisted of a familiarization phase and a test phase. During the familiarization phase infants listened to the continuous familiarization speech stream, which was played independently of infants' looking behavior. Infants also gained experience with the visual attention getters, which unlike the sounds were presented contingently upon infants' looking behavior (see below). After the end of the familiarization phase, infants immediately went on to the test phase. In the test phase, both the sound and the visual stimulus were contingent upon infants' looking behavior. Both familiarization and test trials started with the presentation of the central attention getter on the front screen. Once infants reliably fixated on it, the central attention getter was extinguished and one of the side attention getters was turned on (sides were randomized and counterbalanced within and across infants). Once the infant reliably fixated on the blinking side screen, as indicated by a head turn of at least 30^{o} to that side, a sound stimulus started to play from the loudspeaker placed below the corresponding side screen during the test phrase. The trial lasted until the infant turned away for more than a predefined look-away criterion (2 sec) or until the end of the sound file (21 sec). A new trial was then presented. During the study, an experimenter located outside the testing booth and blind to the stimuli being presented, monitored infants' looking

behaviour and operated the stimulus presentation software (PsyScope version X B55 run on a Mac OS X, version 10.10.5). Experimental sessions were recorded, and the videos were analysed offline to measure infants' looking times by two experienced coders.
The new experimental box was created together with Maria Clemencia Ortiz and Enikö Ladányi, who also collected preliminary data of experiment 3.

Figure 3.1. *The HPP setup using in the current study*

3.3.1.1 Apparatus and stimuli

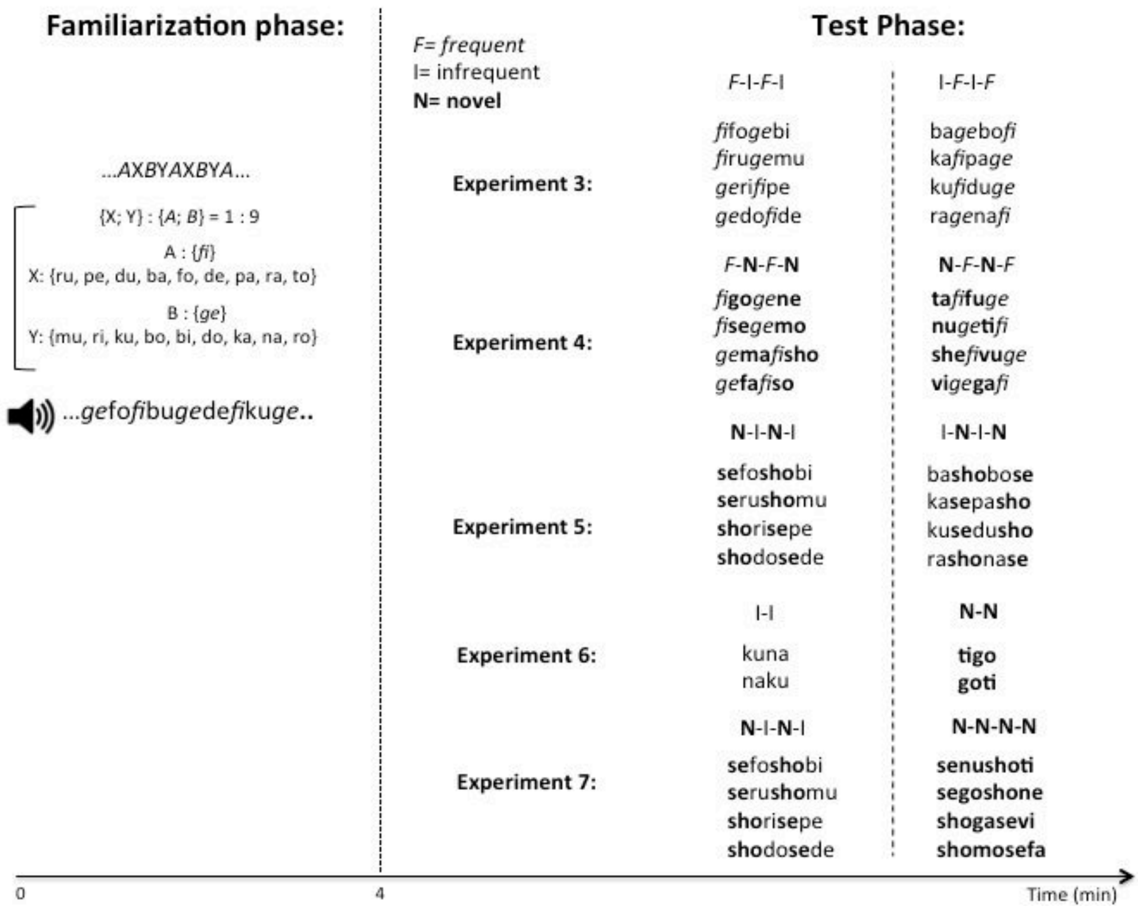
An artificial grammar was created following Gervain and Werker (2013). During familiarization a 3-minute 48-second long speech stream was played with alternating frequent and infrequent words concatenated without pauses. The grammar consisted of a four-syllable-long basic structure (AXBY), where each unit was realized as a consonant-vowel (CV) syllable. In this structure A and B units were frequent words (mimicking function words), whereas X and Y were infrequent words (mimicking content words), because the A and B categories had one token each (A: *fi*; B: *ge*), while the X and Y categories contained nine tokens (X: ru, pe, du, ba, fo, de, pa, ra, to; Y: mu, ri, ku, bo, bi, do, ka, na, ro), making individual X and Y tokens nine times less frequent than A and B tokens. All the syllables respected French phonotactic, but were non-words in the French infant vocabulary.

The familiarization stream, identical across all five experiments, was synthesized using a text-to-speech synthesis software (MBROLA, voice f4) (Dutoit, 1996), with a pitch of 200Hz (corresponding to the fundamental frequency of female voices) and a phoneme duration of 120ms. The stream thus provided no prosodic cue to its structure. Furthermore, the initial and final 15sec of the stream were ramped in amplitude, suppressing information about the exact beginning and end of the stream. As a result, the structure of the stream was ambiguous between a frequent word initial (AXBY or BYAX) and a frequent word final (XBYA or YAXB) parse. During the test phase, respectively, 8 (experiments 3, 4, 5 & 7) or 4 items (experiment 6) were presented. The full list of CV syllables and test items used are presented in detail below (**Figure 3.2**).

Parents of all participating infants gave written informed consent prior to participation. All experiments were approved by the ethics boards of the institutions involved (CERES of the Université Paris Descartes) CER-Paris Descartes, approval nr. 2016/32. Parents were also asked to fill in a questionnaire regarding infants' developmental history and language exposure.

All infants were full-term, with no history of hearing dysfunction or visual problems, no recent occurrences of ear infection and no family history of congenital hearing loss. French was the major language spoken in these families (<20% exposure to other languages). Information about family risk for language impairment was assessed through the parental questionnaire administered during the visit. Language risk was defined as the presence of a first-degree family member (parent and/or sibling) who was diagnosed with dyslexia or other speech impairments.

Figure 3.2. *Artificial grammar task used in experiments 3, 4, 5, 6 & 7*



3.3.2 Experiment 3

3.3.2.1 Participants

Thirty (15 girls and 15 boys) 8-month-old (mean age 8 months and 4 days, range 7.5- 9 months) French infants took part in experiment 3. Among these 30 infants, nine were not included in the final data analysis, because of technical problems (1), because French was not the main language spoken in the home (2), because they had too short (shorter than 960msec, the duration of a test item) or too long (longer than 21msec, the maximal duration of a trial) looking times in more than four trials (4) or for fussiness and crying (2). A final sample of 21 participants were entered into the analysis.

3.3.2.2 Stimuli

Infants were exposed to the familiarization stream identical to the one used in Gervain et al. (2008). They were then tested on sequences taken from the familiarization stream following

Gervain et al. (2008). Infants were presented with eight four-syllabic chunks from the stream; half of them started with a frequent word, the other half with an infrequent word (e.g. *fi*fo*ge*bi vs. ba*ge*bo*fi*, for further details see Figure 3.2 and **Table 3.2**). Test trials had a total duration of 21sec, as a single item was repeated 16 times with pauses of 500ms. A trial was stopped if the infant looked away for more than 2 seconds. Thus, trial duration was contingent upon infant looking time. French being a functor-initial language, here we predicted that infants would show a preference for the frequent-word initial sequences, following previous results with Italian and Japanese infants (Gervain et al., 2008).

Table 3.2. *Test items used in experiment 3. F: frequent items;* I: infrequent items

F-I-*F*-I	I-*F*-I-*F*
*fi*fo*ge*bi	ba*ge*bo*fi*
*fi*ru*ge*mu	ka*fi*pa*ge*
*ge*ri*fi*pe	ku*fi*du*ge*
*ge*do*fi*de	ra*ge*na*fi*

3.3.3 Experiment 4

For a second group of infants, we replaced the infrequent words in the test items with novel ones. If infants expected infrequent words to be content words, thus belonging to open classes, they should maintain their frequent word initial preference despite the novel words, as the frequent "functors" providing the structural skeleton of the sequences remained in place. The novel words in the test items could thus be categorized as "infrequent" despite the fact that no frequency information was available about them given their position with respect to the unchanged frequent words.

This experiment was designed by Carline Bernard under Judit Gervain's supervision. Data was collected by Carline Bernard and Lisa Jacquey.

3.3.3.1 Participants

Thirty-three (16 girls and 17 boys) 8-month-old (mean age 8 months and 20 days, range 8- 9 months) French infants took part in experiment 4. Among these 33 infants, twelve were not included in the final data analysis, because they had too short or too long looking times (4) or for fussiness and crying (8). A final sample of 21 participants were entered into the analysis.

3.3.3.2 Stimuli

The procedure and the familiarization stream were identical to the ones used in experiment 3. During the test phase the infrequent words were replaced by new CV syllables (X': **go**, **so**, **se**, **sho**, **nu**, **vi**, **fu**, **vu**; Y': **ne**, **fa**, **mo**, **ma**, **ti**, **ga**, **ta**, **she**) (further details in Figure 3.2 and **Table 3.3**).

Table 3.3. *Test items used in experiment 4. F: frequent items;* **N**: **novel items**

F-**N**-*F*-**N**	**N**-*F*-**N**-*F*
*fi***go***ge***ne**	**ta***fi***fu***ge*
*fi***se***ge***mo**	**nu***ge***ti***fi*
*ge***ma***fi***sho**	**she***fi***vu***ge*
*ge***fa***fi***so**	**vi***ge***ga***fi*

3.3.4 Experiment 5

For a third group of infants, we replaced the frequent words with novel ones. We expected this manipulation to disrupt infants' preference, as they could no longer rely on the frequent words as structural anchors. It was still possible to categorize the novel words as "frequent" given their position, but if infants treated infrequent words as content words, they may have been less likely to rely them for structural information when making a word order choice. In experiment 4 and 5, it was the position and not the frequency information that allowed generalization (or not) to novel items in the test phase.

3.3.4.1 Participants

Twenty-seven (10 girls and 17 boys) 8-month-old (mean age 8 months and 20 days, (range 8-9 months) French infants took part in experiment 5. Among these 27 infants, 6 were not included in the final data analysis, because they had too short or too long looking times (3) or for fussiness and crying (3). A final sample of 21 participants entered in the analysis.

3.3.4.2 Stimuli

The procedure and the familiarization stream were identical to the ones used in experiments 3 and 4. During the test phase the frequent words were replaced by new CV syllables (A': **se**; B': **sho**) (further details in Figure 3.2 and **Table 3.4**).

Table 3.4. *Test items used in experiment 5.* I: infrequent items; **N**: **novel items**

N-I-N-I	I-N-I-N
***se**fo**sho**bi*	ba***sho**bo**se***
***se**ru**sho**mu*	ka***se**pa**sho***
***sho**ri**se**pe*	ku***se**du**sho***
***sho**do**se**de*	ra***sho**na**se***

3.3.5 Experiment 6

As a control, experiment 6 was carried out to investigate whether infants, actually remembered the infrequent words from the familiarization stream. This tested whether any significant difference between experiments 4 and 5 really arose from differences in how infants categorized frequent and infrequent elements, and not simply because of a failure to remember the infrequent ones.

3.3.5.1 Participants

Thirty (13 girls and 17 boys) 8-month-old (mean age 8 months and 16 days, (range 8-9 months) French infants took part in experiment 6. Among these 30 infants, five were not included in the final data analysis, because they had too short or too long looking times (1) or for fussiness and crying (4). Additionally, seven infants were analysed separately because of a family risk/history of language impairment. A final sample of 18 participants were entered into the analysis.

3.3.5.2 Stimuli

The procedure and the familiarization stream were identical to the ones used in experiments 3-5. During the test phase infants were presented with two familiar infrequent items and two novel items (further details in Figure 3.2 and **Table 3.5**). If the infants were able to remember the infrequent items, they should show different looking times for one of the two types of test items. A novelty preference, i.e. longer looking times to the novel items, would be a stronger indicator of recognition/memory, as it is believed to arise when the processing of the familiarization material is completed during the familiarization phase itself.

Table 3.5. *Test items used in experiment 6.* I: infrequent items; **N**: **novel items**

I-I	N-N
kuna	**tigo**
naku	**goti**

3.3.6 Experiment 7

Finally, a fifth group of infants was tested to investigate whether infants encode the position of infrequent words at all. We presented infants with test items where frequent words were replaced by novel words in the initial position (e.g. **se**fo**sho**bi, as in the **N**-I-**N**-I test items in experiment 5) as well as with items in which both frequent and infrequent words were replaced by novel ones (e.g. **senushoti**). These latter items carry neither frequency, nor positional information. If infants under these conditions show a preference for the items in which infrequent words were in place and were in the native-like final position, then that would suggest that even if they do not readily rely on infrequent words as structural anchors, they are nevertheless able to encode their position. If they don't use infrequent words to encode position, a novelty preference is expected just like in experiment 6.

3.3.6.1 Participants

Thirty (15 girls and 15 boys) 8-month-old (mean age 8 months and 6 days, range 7.5-9 months) French infants took part in experiment 7. Among these 30 infants, ten were not included in the final data analysis because of technical problems (1), because they had too short or too long looking times in more than four trials (7) or for fussiness and crying (2). A final sample of 20 participants was entered into the analysis

3.3.6.2 Stimuli

The procedure and the familiarization stream were identical to the ones used in experiments 3-6. During the test phase infants were presented with two types of test items. In one, the two frequent words were replaced by new CV syllables (A': **se**; B': **sho**), whereas in the second both the infrequent words (X': **go**, **nu**, **vi**, **fa**; Y': **ne**, **ti**, **ga**, **mo**) and the frequent words (A': **se**; B': **sho**) were replaced by new CV syllables (further details in Figure 3.2 and **Table 3.6**).

Table 3.6. *Test items used in experiment 7.* I: infrequent items; **N**: **novel items**

N-I-N-I	N-N-N-N
***se**fo**sho**bi*	**senushoti**
***se**ru**sho**mu*	**segoshone**
***sho**ri**se**pe*	**shogasevi**
***sho**do**se**de*	**shomosefa**

3.4 Data analysis

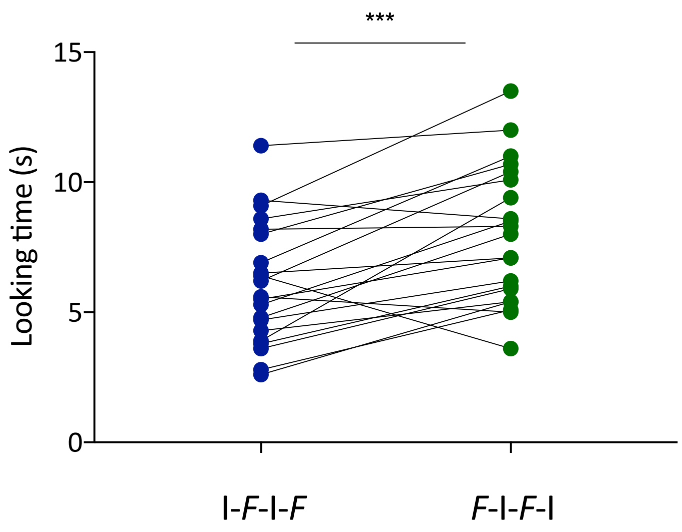
The videos were analysed offline with the software PsyCode: (http://psy.ck.sissa.it/psycode/psycode.html). For each experiment, we averaged looking times across all trials of the same condition after the offline coding of the videos. One blind coder coded all the videos. Additionally, a second blind coder coded 18% of them. The correlation between the two coders was r = 0.85. We performed paired samples t-tests with equal variance not assumed to compare the two test item types in each experiment. To directly compare our key experimental manipulations, we also submitted the looking time data to an analysis of variance (ANOVA) with Category Replaced (infrequent/ frequent) as a between-subject factor and Test Item (frequent position initial/frequent position final) as a within-subject factor.

3.5. Results

3.5.1 Experiment 3

Results of experiment 3 are shown in **Figure 3.3.** The average looking time across all trials of the same type [frequent initial vs. frequent final items] was calculated after offline analysis and a paired samples t-test (with equal variance not assumed) between the two test item types was performed. Infants in experiment 3 demonstrated the predicted frequent word initial preference (*F*-I-*F*-I looking times: M = 8.0s; SD = 2.6; I-*F*-I-*F* looking times: M = 6.08s; SD = 2.3; $t(20) = 4.41$; $p = 0.0003$; $d = 0.77$; power $(1-\beta) = 0.91$), corresponding to French word order. At the individual level, 18 out of 21 infants listened longer to the frequent initial items than the frequent final items. Binomial tests indicate that this proportion is significantly higher than expected by chance ($p = 0.0015$). **Figure 3.3** reports the distribution of looking times at the individual level. This result establishes that 8-month-old French-learning infants do indeed represent the basic word order of their native language in terms of the relative order of frequent and infrequent words. This knowledge is robust even at the individual level, as almost all infants show the predicted preference.

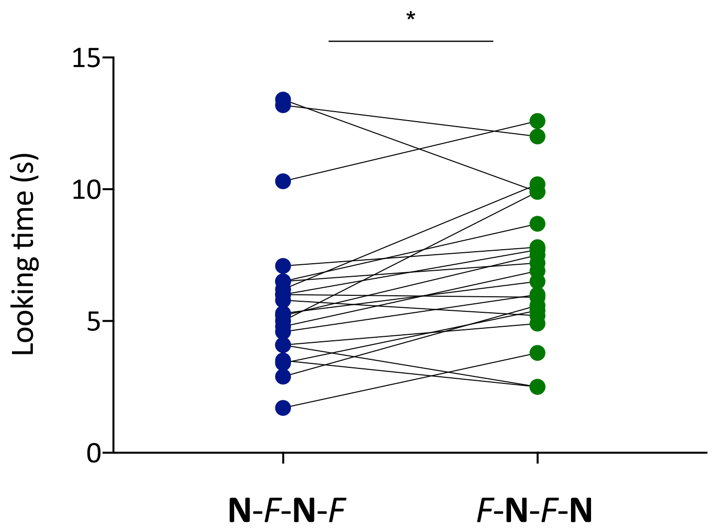
Figure 3.3. *Plot of the individual variability in looking times in experiment 3.*



3.5.2 Experiment 4

Results of experiment 4 are shown in **Figure 3.4.** Crucially, the frequent initial preference was maintained in experiment 4, even though the infrequent words were replaced by novel ones in the test items (*F*-**N**-*F*-**N** looking times: M = 7.1s; SD = 2.7; **N**-*F*-**N**-*F* looking times: M = 6.0s; SD = 3.0; t (20) = 2.592, p = 0.01; d = 0.38; power (1-β) = 0.38). 15 out of 21 infants have listened longer to the frequent initial items. Binomial tests indicate that this proportion is marginally higher than expected by chance (p = 0.07). This finding shows that infants treat the infrequent category as an open class, which is a key feature of content word categories in natural language. Given the novel and thus surprising nature of the manipulation in this experiment, results at the individual level are somewhat weaker than in experiment 3.

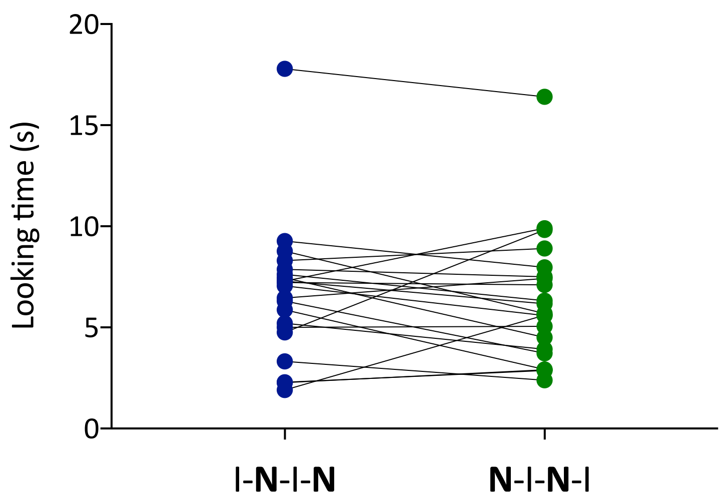
Figure 3.4. *Plot of the individual variability in looking times in experiment 4.*



3.5.3 Experiment 5

Results are presented in **Figure 3.5**. Infants did not showed a word order preference (**N**-I-**N**-I looking time M = 6.32s: SD = 3.23; I-**N**-I-**N** looking time M = 6.64s; SD = 3.34; t (20) = 0.677, p = 0.51; d = 0.12; power (1-β) = 0.09), confirming that they do not accept novel frequent words, treating this category as a closed class. Here only 8 out of 21 infants listened longer to the frequent initial test items. Binomial tests indicate that this proportion is not significantly different from chance (p = 0.383). This suggests that they expect frequent words to behave like functors in natural language, constituting a closed class.

Figure 3.5. *Plot of the individual variability in looking times in experiment 5.*



Overall, experiments 4 and 5 provide key evidence that infants treat frequent and infrequent words differently in terms of their replaceability. Thus, it seems that the categorization of function and content words, with their respective characteristics, is present from the earliest stages of language acquisition.

To directly test whether infants behaved differently in the baseline condition (experiment 3), when infrequent (experiment 4) and frequent (experiment 5) words were replaced, we conducted a repeated measures ANOVA with Category Replaced (none/infrequent/frequent) as a between-subject factor and Test Item (frequent position initial/frequent position final) as a within-subject factor. We obtained a significant Test Item effect [$F(2,60) = 12.35$, $p < 0.001$], due to longer looking times to frequent-initial items than to frequent final items, and a significant Category Replaced X Test Item interaction [$F(2,60) = 6.472$ ($p < 0.005$)] due to significantly longer looking times for frequent position initial items than frequent position final item in experiment 3 (Scheffe's post hoc test: $p < 0.001$), experiment 4 (Scheffe's post hoc test: $p = 0.051$) but not in experiment 5 (Scheffe's post hoc test: $p = 0.77$, Ns.) as well as longer looking times to frequent initial items in experiment 3 than in experiment 5 (Scheffe's post hoc test: $p = 0.002$). No other pairwise comparison was significant, including the looking times to frequent initial items in experiments 3 and 4 (Scheffe's post hoc test: $p = 0.14$, Ns.).

3.5.4 Experiment 6

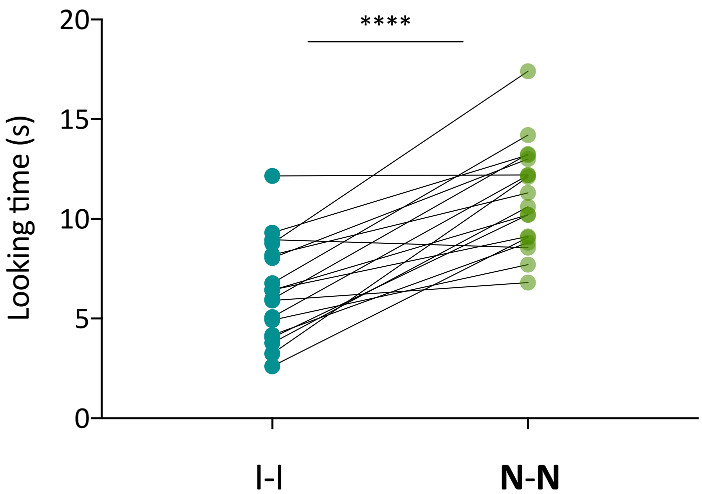
Importantly, in the familiarization stream, frequent words were nine times more frequent than the infrequent ones. It is thus possible that infants no longer showed a word order preference when the frequent words were replaced (experiment 5) not because they couldn't parse the underlying structure, but simply because the frequent words were the only words they remembered from the familiarization stream. Thus their replacement suppressed all relevant information. If that was the case, replacing the infrequent words would indeed make no difference to infants, because they do not recall them and hence might fail to notice the replacement. To exclude this alternative explanation, in experiment 6, we presented infants with pairs of infrequent words from the stream as well as with entirely novel words.

3.5.4.1 Results: typically developing infants

Results are presented in **Figure 3.6**. Infants showed a preference for the novel items (novel items looking time (**N-N)**: M = 11.1s; SD = 2.6; familiar items looking time (I-I): M = 6.38s; SD = 2.5; t (17) = 7.084, $p < 0.0001$; d = 1.84; power (1-β) = 1). At the individual level, 17 out of 18 infants listened longer to the novel items than to the familiar ones. Binomial tests indicate that this proportion is very highly significantly different from chance (p = .0001). This novelty preference is an indicator that infants were more familiar with the infrequent words that appeared in the stream than with the novel items. Their differential looking behaviour in experiments 4 and 5 cannot thus be attributed to memory limitations in encoding the words of the stream.

In the previous experiments, we obtained a familiarity preference, while here a novelty preference. This is because the tasks were quite different. In the previous experiments, we presented infants with two types of test items: one that was representative of their native word order and one that wasn't. Here instead we used items that were familiar and items that were completely new. Infants remembered and recognized the infrequent words from the familiarization phase and looked longer at novel items, as they might have been more interested in them.

Figure 3.6. *Plot of the individual variability in looking times in experiment 6.*

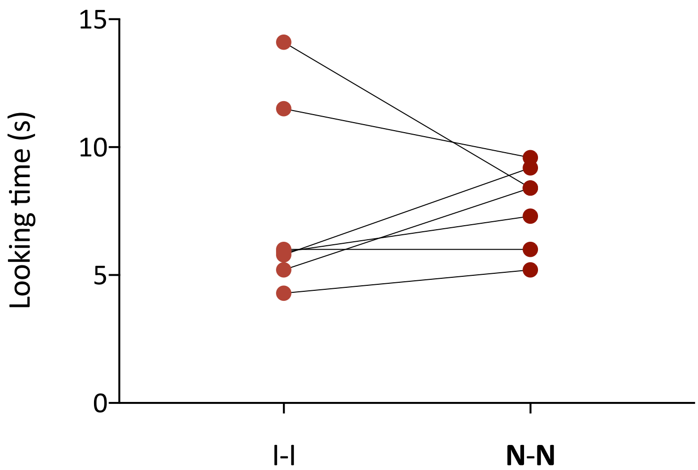


3.5.4.2 Results: infants at risk

Seven infants in the group were analysed separately because of a family history of language impairment. As discussed in the introduction, one of the general aims of this study is to investigate whether differences in behaviour between typical and atypical participants are present and predict later language outcomes.

At the group-level, contrarily to the controls, infants at risk showed no significant difference in the mean looking time between the two test items (novel items looking time (**N-N)**: M = 7.7s; SD = 1.6; familiar items looking time (I-I): M = 7.5s; SD = 3.7; $t(6) = 0.155$, $p = 0.88$). In order to analyse the individual differences, the distribution of the individual looking times within the group was plotted (**Figure 3.7**). While the mean looking time for the novel items was quite homogeneous between participants, a huge difference was observed for the familiar infrequent items, suggesting that not all the infants were able to remember the infrequent items in the same way.

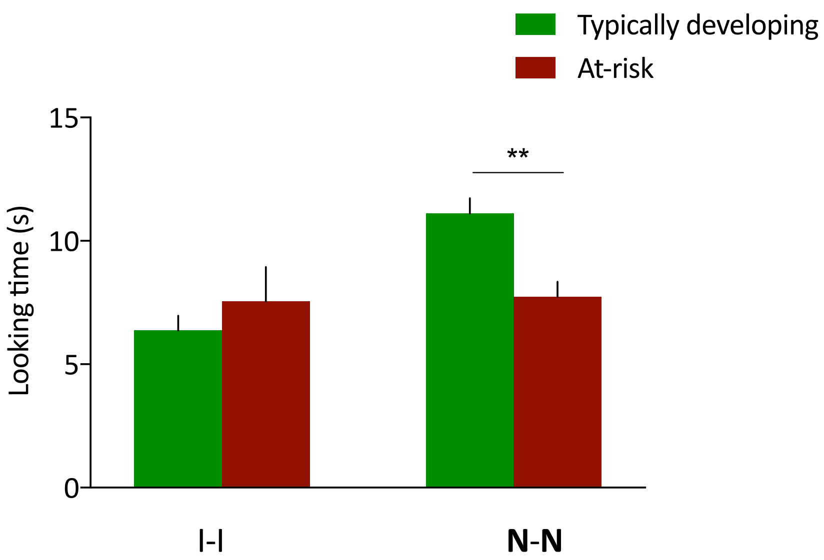
Figure 3.7. *Plot of the individual variability of the looking time for novel vs. familiar test items within the group of infants at risk.*



3.5.4.3 Comparison between the two groups

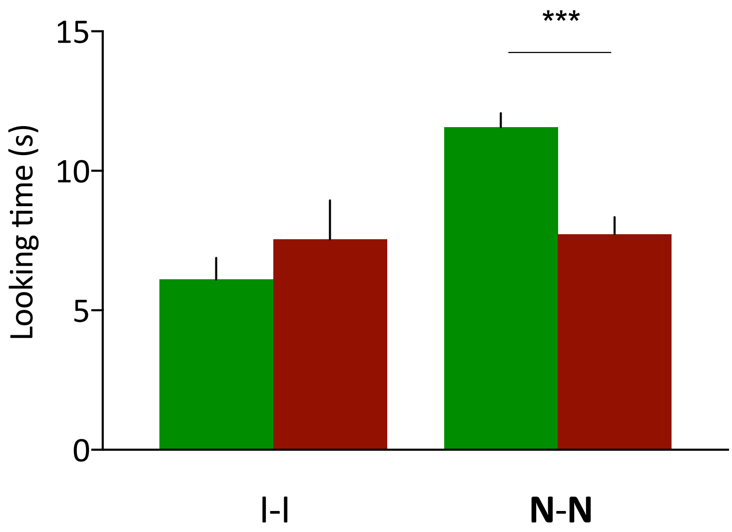
An ANOVA with Group (typically developing vs. at-risk) as a between-subject factor and Test Item (novel/familiar) as a within-subject factor with looking time as the dependent variable was run. We obtained a significant effect of Test Items [$F(1,23) = 14,23$, $p = 0.001$] and a significant Group X Test Item interaction [$F(1,23) = 12,17$, $p = 0.002$] due to a greater difference between looking times to the familiar and infrequent items in the typical group than in the at-risk group.

Figure 3.8. *Comparison between typically developing (n =18) and at-risk (n =7) groups. The x-axis shows the different experimental groups and conditions. The Y-axis shows the looking time in seconds. Error bars represent the s.e. of the mean.*



It needs to be noted, however, that the two groups had different sample sizes (typically developing: n= 18: at risk: n= 7). Thus, for a more balanced comparison, seven infants were randomly selected from the typical group. In line with the whole group comparison an ANOVA with Group (typically developing vs. at-risk) as a between-subject factor and Test Item (novel/familiar) as a within-subject factor with looking time as the dependent variable was run obtaining a significant effect of Test Items [$F(1,12) = 16{,}80$, $p = 0.001$] and a significant Group X Test Item interaction [$F(1,12) = 14{,}66$, $p = 0.002$] due to a greater difference between looking times to the familiar and infrequent items in the typical group than in the at-risk group. The results (**Figure 3.9**) are the same as for the whole group comparison.

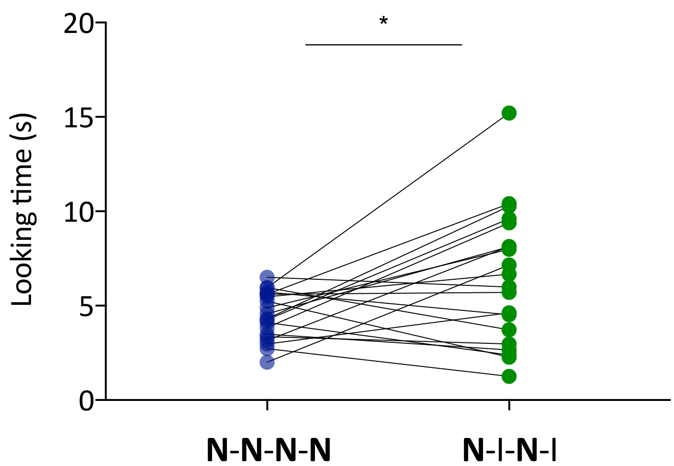
Figure 3.9. *Comparison between typically developing (n =7, in green) and at-risk (n =7, in red) groups. The x-axis shows the different experimental groups and conditions. The Y-axis shows the looking time in seconds. Error bars represent the s.e. of the mean.*



3.5.5 Experiment 7

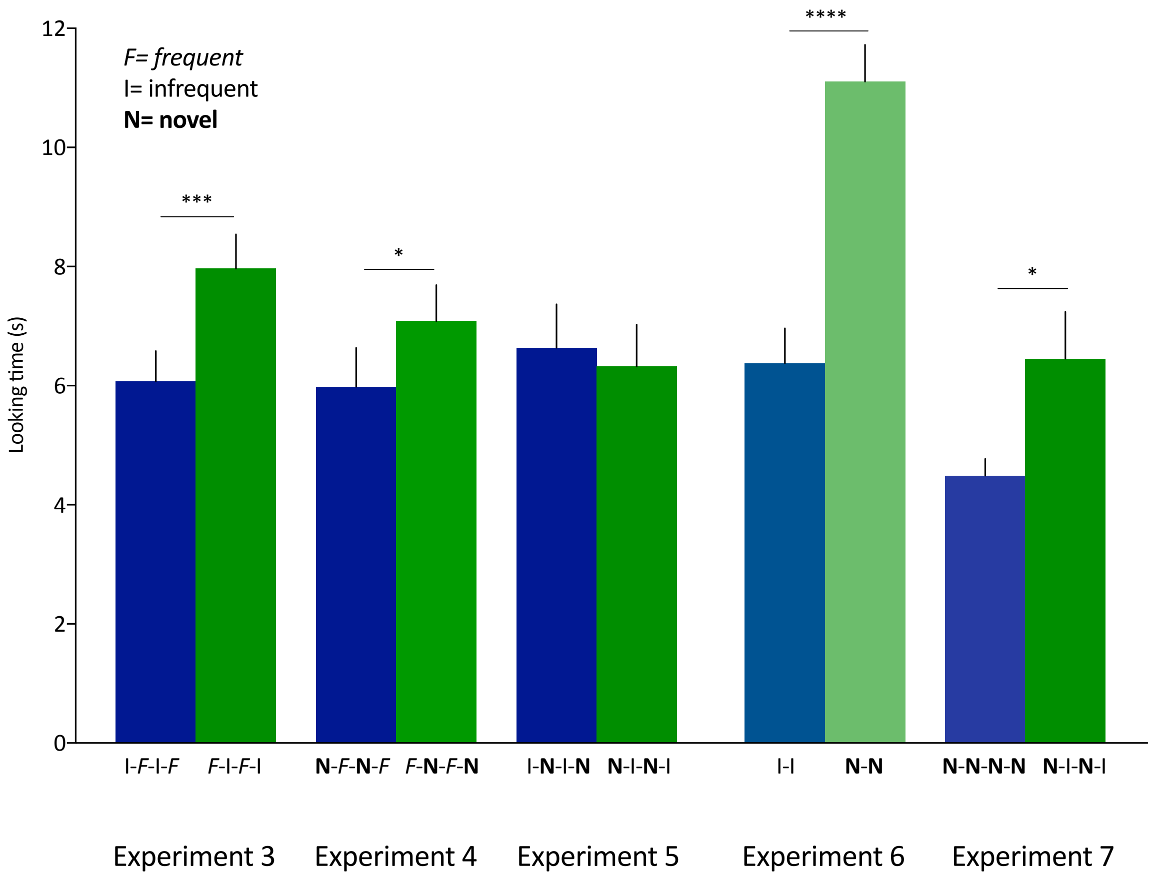
Results are presented in **Figure 3.10**. Infants here showed a preference for the items in which infrequent words signalled information about position (**N**-I-**N**-I looking time: M = 6.45 s; SD = 3.5; **N**-**N**-**N**-**N** looking time M = 4.49 s; SD = 1.3; t (19) = 2.55, p = 0.02; d = 0.74; power (1- β) = 0.86). At the individual level, 12 out of 20 infants showed the preference. Binomial tests indicate that this proportion is not different from chance (p = 0.5).

Figure 3.10. *Plot of the individual variability in looking times in experiment 7.*



Experiment 5 and experiment 7 both had **N**-I-**N**-I test sequences. Interestingly, there was no difference in the mean looking times for these test items in the two experiments (unpaired t test, with equal variance not assumed: **N**-I-**N**-I, experiment 5: $M = 6.32$; **N**-I-**N**-I, experiment 7: $M = 6.45$; $t(38.28) = 0.123$; $p = 0.9$).

Figure 3.11. *Looking time results for experiments 3, 4, 5, 6 & 7. The x-axis shows the different experimental groups. The Y-axis shows the looking time in seconds. Error bars represent the s.e. of the mean.*



3.6 Discussion

In five artificial grammar learning experiments, we have found that infants treat frequent words as belonging to closed classes, and infrequent words as belonging to open classes, and they map the relative order of these categories onto the order of functors and content words in their native language. These findings provide the earliest evidence that infants as young as 8 months of age already use word frequency as a cue to lexical categories, which they combine functionally to build rudimentary representations of grammar. This implies that the acquisition of grammar starts at the preverbal stage of language learning, in parallel with the development of native phonology and the lexicon.

This early acquisition can readily be accounted for within a bootstrapping framework of language acquisition. In this framework, our results suggest that the differential frequency distribution of functors and content words is a cue that infants as young as 8 months of age can rely on to bootstrap the basic word order of their native language.

Frequency-based word order bootstrapping relies on the language universal distinction between function and content word (Fukui, 1986: Abney, 1987). The ability to identify functors is a crucial first step for infants on their way to their native grammar. Functors are highly frequent in language, while content words tend to occur less frequently. It has, therefore, been hypothesized that infants rely on frequency of occurrence as a particularly useful cue to establish these two basic lexical categories. However, direct evidence for this assumption was still missing. It was thus unknown how and at what age infants first recognize and use these basic building blocks of language to parse the input into grammatically relevant patterns.

It has been proposed that adults and older children track functors in the input and use them as anchors, encoding the position of other words in relation to them (Valian & Coulson, 1988; Morgan, Meier and Newport, 1987; Bernard and Gervain, 2012). Indeed, monolingual and bilingual adult speakers prefer the relative order of functors and content words in an artificial grammar that is coherent with the basic word order of their native language (Gervain et al., 2013; de la Cruiz et al., 2014). Similarly, 8-month-old infants exposed to languages with opposite word orders, e.g. functor-initial Italian and functor-final Japanese, show opposite preferences for word order in an artificial grammar task. Italian infants prefer sequences starting with a frequent word, while Japanese infants prefer sequences starting with an infrequent word, mirroring the word orders of these two languages (Gervain et al., 2008). As the difference in frequency between functors and content words is a common design feature of all natural languages, the frequency-based bootstrapping strategy of lexical categories may operate universally across languages and could provide a powerful tool for young infants to break into language.

The ability to track words on the basis of frequency has also been found in non-human animals such as rats (Toro et al., 2016). This suggests that frequency-based categorization is a general mechanism shared across species, and thus available to very young infants. Crucially, however, rats always show a preference for frequent-initial sequences. This behaviour is different from that of infants, who prefer the order that matches the word order found in their native language, i.e. frequent-initial for French and Italian infants, but frequent-final for Japanese infants. This implies that frequency-based lexical categorization interacts with language experience and feeds into linguistic representations, constituting a strong bootstrapping strategy (Gervain et al., 2008; Gervain and Werker, 2013).

In apparent contradiction with our hypothesis, functors are generally produced later than content words. Crucially, however, while infants produce functors relatively late, in perception they are sensitive to them much earlier. This may be due to the difference between the phonological make up of the two classes. Functors are universally phonologically more reduced than content words, e.g. depending on the language, they may not carry stress, have a simpler syllable structure than content words, often contain reduced vowels etc. (Morgan, Shi & Allopenna, 1996). The specific features in which they differ from content words varies across languages, but they are always more minimal. In French, for instance, grammatical functors cliticize onto their content words (e.g. *je t'aime* /ʒ(ə)**'tɛ**m(ə)/ I you.acc love 'I love you'), they are typically shorter and have simpler syllable structure than content words. Importantly, the omission of functors in early production is therefore attributable to their phonological weakness and not to infants' inability to perceive or represent them (Gerken et al., 1990). Generally, infants produce functors later than content words due to production constraints attributable to their phonological minimality (e.g. tendency to omit unstressed rather than stressed word units) (Demuth, 1994; Gerken & McIntosh, 1993). The fact that infants do not produce functors early is thus not incompatible with our result that at 8 months they can already track them in the input and use them as structural anchors.

In the present study, we have shown that infants as young as 8 months of age already known some of the most general properties of lexical categories and their native language grammar. Since at 8 months, they do not yet have a sizeable lexicon, this knowledge is most likely not item based. Rather, bootstrapping appears to be a learning mechanism independent of vocabulary learning. Overall, these findings are in line with the existence of abstract syntactic and lexical knowledge, supporting the hypothesis that during the development genetically endowed general language principles interact with language-specific experience and structural knowledge.

Importantly, a sub-group of seven infants with family history of language impairments participated in experiment 6. One of the general aims of this work was to provide evidence for early differences between typical and atypical populations. Thus, in experiment 6, a parallel analysis was conducted. The group of typically developing infants performed as expected, showing a preference for the novel items. This suggests that they remembered infrequent words in the stream. Infants at-risk, by contrast, didn't show a preference for any of the test items suggesting a weaker encoding, recall or recognition of the infrequent items. These abilities are all crucial during language acquisition. One factor that might have played a role is the time of exposure, i.e. the amount of familiarization that was provided. Infants with limitations in recognition and recall are likely to need more repetitions or more time in order to retrieve the information. In fact how quickly an infant makes this "familiarity-to-novelty shift" depends on their processing speed and the complexity of the stimuli (Hunter and Ames, 1988). Our finding is perfectly in line with the general perspective that infants at risk have atypical attention and memory abilities and perhaps need more exposure time or more trials in order to achieve the same proficiency as their peers. Nevertheless, it was less clear whether infants at-risk were poorly in their ability to encode the familiar infrequent items during the familiarization phase or they, instead, had less ability to recall them during the test phase. Moreover, some of the at-risk infants showed a novelty preference but the performance at the group level was overall much variables compared to typical peers. Therefore, further investigation, e.g. collecting expressive and receptive vocabulary data, might help to understand whether these early differences lead to different/delayed linguistic outcomes later in the development.

Overall, our findings present the first evidence that 8-month-old infants are already able to extract frequency and positional information from language input to establish the basic lexical categories of function and content words, and use these to build rudimentary representations of grammar. The distinction between function and content words is a universal feature of human language, which infants need to learn early in development. Frequency-based bootstrapping, the mechanism we have uncovered here, provides an account of how they might achieve this early and efficiently.

Chapter 4:

General Discussion

The present work has investigated the perception of surface cues present in the speech input and how this perception shapes later language acquisition. It has proposed that bootstrapping mechanisms play a considerable role in helping infants learn abstract structural properties of their native language by using perceptually available surface cues from the input (Morgan & Demuth, 1996). Moreover, it has investigated whether these bootstrapping mechanisms vary across typically developing infants as well as between typical and atypical infants, potentially providing early behavioural markers of risk for language delay.

Auditory thresholds and their relation with later linguistic development were explored in chapter 2. Two psychoacoustic experiments tested whether infants who are better processors of rapidly presented auditory stimuli (both linguistic and non-linguistic), have larger vocabulary later during development. Moreover, as several external factors might contribute to infants' language development, correlational patterns between general cognitive measures and auditory processing abilities were investigated together with the existence of a link between these cognitive skills and later cognitive and linguistic development. Specifically, in experiment 1 and 2 infants performed an auditory discrimination threshold task (using the rapid auditory processing paradigm), followed by a visual novelty detection task used as a control for general cognitive skills (following Benasich and Tallal, 1996). The auditory discrimination threshold was evaluated in two cohorts using non-linguistic (tones) and linguistic sounds (syllables). Subsequent measures of vocabulary abilities (CDIs) were collected at 12-14-18 and 24 months and a cognitive test (The Mullen Scale of Early Learning) was performed at 18-20 months as an additional measure for early general processing abilities.

Bootstrapping mechanisms that support abstract structural learning were investigated in chapter 3. In particular, word frequency was tested as a cue to the basic lexical categories of function and content words and to their relative position in the native word order. In experiments 3-7, artificial grammar learning experiments were run providing evidence that frequency-based bootstrapping supports lexical categorization of functors and content words and the relative acquisition of native word order.

Importantly, since the literature suggests that dyslexic adults and children have impaired auditory discrimination, comparative analyses were run between typically developing infants and infants at-risk for developmental dyslexia (DD). This work reveals important differences in the general processing abilities between typical and atypical populations. Behavioural markers of early risk were identified. Infants at-risk for DD presented slower processing efficiency in both the auditory and the visual modalities, which was correlated with a lower score in productive vocabulary both at 14 and 18 months.

4.1 Typical populations

Bootstrapping approaches assume that perceptual learning principles, guiding infants' knowledge of grammatical structure, rely on surface cues available in the speech input (Morgan & Demuth, 1996). Understanding the relation between early perceptual abilities and later language outcomes might help us determine the extent to which early processing measures are language-specific. However, bootstrapping mechanisms are heuristic, often apply differently in different languages and rely on surface cues of different nature (acoustic, statistical etc.). Therefore, the degree and the efficiency at which infants process these surface cues may differ across individuals resulting in different developmental trajectories or strategies.

While the relation between variability in early speech processing abilities and later language learning is a central issue in atypical development, the majority of studies with typically developing infants assume that the development of the linguistic skills is homogeneous in this population. Nevertheless, even if infants go through the same sequence of gross developmental milestones, substantial variability is present in the timing, quality and efficiency of early linguistic abilities in typically developing participants. For instance, variation in early vocabulary growth might be potentially related to the speed and the efficiency of early auditory processing in typical population (e.g. Benasich and Tallal, 2002). But how is the early perception of the surface features of speech correlated with language development?

Speech is, first, a fast changing signal and these fast changes have to be processed accurately in order to understand language in real time. It therefore requires efficient processing abilities; otherwise incoming information might be lost. The ability to process fast changes is fundamental to correctly perceiving speech. Infants who are better early processors may develop more robust phonological and lexical representations (e.g. Benasich & Leevers, 2002; Benasich & Tallal, 2002; Fitch et al., 2001; Farmer & Klein, 1995).

A few studies to date have investigated individual variability in auditory threshold measures in typical populations. Trehub and Henderson (1996) investigated the relation between early auditory abilities (at 6 and 12 months) and later language outcomes. Infants who performed better at early stages obtained larger expressive vocabulary between 16 and

29 months. This confirms that auditory processing skills may underlie differences in language abilities in typically developing infants.

The current study could only partially answer this question, as the results observed in the psychoacoustic tasks were not fully reliable, most probably due to the methodological difficulties of the experimental procedure, underestimating infants' rapid auditory thresholds. As a result, psychoacoustic performance was not predictive of later language outcomes either for tones or syllables, but this null result likely reflects methodological limitations and provides no information about the existence of a connection between rapid auditory processing and language development.

Relatedly, the efficiency with which infants processed the low-level acoustic cues was highly variable. In both cohorts, individual performance showed great between-listener variability. Consequently, participants were split into sub-groups based on the shape of the adaptive procedure during the test phase. The identification of different profiles contributes to a better understanding of individual variation in auditory processing abilities and highlights the high attentional and cognitive demands of this task. Three groups were established: the *good*, the *poor* and the *inconsistent/unstable* performers (based on Moore et al., 2008). Generally, infants who were better processors of rapid acoustic information were those who also showed more stability in general attentional level.

In the group of the *good* performers, infants learned the procedure faster and achieved the acoustic threshold quickly. However, even in this group, the mean auditory threshold was much higher than expected. The *poor* performers learned the procedure efficiently and had a similar pattern to the *good* profile, except that the individual thresholds were higher. Even if the absolute values of the thresholds are not fully reliable, it is still likely that a genuine difference exists between good and poor performers as they show the expected adaptive staircase profile.

Lastly, for the *unstable/inconsistent* profile, the acoustic thresholds, even if higher, did not significantly differ from those of the *good* performers. Nevertheless, infants in this group exhibited distant reversal points and several fluctuations in attention during the task (e.g. presence of null responses and repetitive mistakes).

Relevantly, between the tone and syllable cohorts the *unstable/inconsistent* performers showed a different behaviour. In the tone cohort, unstable performance was not attributable to poorer acoustic abilities per se, but to variable attention during the task. In the syllable cohort, by contrast, infants with an *unstable* profile not only exhibited highly fluctuating attention but also auditory thresholds higher than *good* performers for both tones and syllables. Therefore, perhaps, the processing of linguistic sounds might require finer and more stable acoustic and attentional abilities.

The current thesis also investigated whether general attention and cognitive skills were linked to both early acoustic efficiency and to later linguistic outcomes. No direct relation between the psychoacoustic performance and the visual novelty detection task were observed; again most likely due to the insensitivity of the psychoacoustic measures. By contrast, the relationship between general cognitive skills and language outcomes has been found to be strong and stable in our experiments. In particular, the % of novelty effect was predictive of the receptive and expressive vocabulary between 12 and 24 months in the tone cohort. This correlation is in line with previous findings (e.g. Thompson et al, 1991; Rose et al., 2009). Such a strong correlation wasn't observed in the syllable cohort. One potential reason for this may be the unreliability of the parent questionnaire for vocabulary size.

This shows that in addition to language-specific processes, assessing language-external factors might bring important contribution to language acquisition studies, especially those interested in individual variation.

Moreover, the studies also suggested a link in general cognitive skills across the time course of the longitudinal observation. The total score of the Mullen Scale (between 18 and 20 months) was correlated with the % of novelty effect measured in the same infants (syllable cohort) at 9 months.

We also investigated how typical learners may establish a fundamental aspect of native grammar, its basic word order. In experiments 3-7, word frequency was tested as a cue to the basic lexical categories of function and content words and to their relative position in the native word order. These five artificial grammar learning tasks demonstrated that sensitivity to word frequency allows infants to categorize infrequent items as content words belonging to open lexical classes and frequent words as the closed lexical class of function words. Importantly, distinguishing the two lexical categories helped participants map the relative position of frequent and infrequent words onto the native order of functors and content words in real language.

Experiment 3 extended previous findings (Gervain et al., 2008) with Italian and Japanese participants by showing that French monolingual infants also preferred test items that mimic the distribution of function and content words in real language. French being a frequent-initial language, participants significantly preferred test items in which frequent words were placed in initial position.

Experiments 4 and 5, the crucial experimental manipulations, investigated how infants behaved when infrequent words (experiment 4) or frequent words (experiment 5) were replaced by novel words in the test items. Crucially, the frequent-initial preference was maintained when the infrequent words were replaced by novel ones, showing that infants treat the infrequent category as an open class, which is a key feature of content word categories in natural language. Contrarily, this preference was not maintained when the frequent items were replaced (experiment 5), confirming that infants do not accept novel frequent words, treating this category as a closed class. Overall, experiments 4 and 5 suggest that the acquisition of grammar starts before specific lexical items or their respective meanings are acquired.

Experiment 6 showed that despite a null preference in experiment 5, infants remembered the infrequent words in the stream, controlling for a possible memory confound. As a further control, we showed in experiment 7 that if no other information is available, infants can rely on infrequent words to recognize native word order, as long as they don't need to make a word order choice.

For the time being, the results of experiment 3-7 have only been analysed at the group level. Impressionistically, individual looking times show great variability**.** It will be interesting in future studies to explore possible connections with later language outcomes, i.e. to test whether infants who show a stronger preference for the native word order at 8 months have larger vocabularies or higher grammatical competence later.

Our results with typical infants, even if not conclusive, suggest long-term predictability of language growth of the basis of individual processing abilities in the acoustic and visual modalities. Even if correlational studies cannot establish causality, a certain continuity has been observed for some measures.

4.2 Atypical populations

This thesis has revealed important differences in the general processing abilities of atypical participants as compared to their typical peers. The two atypical groups of infants participating in experiments 1-2 (n = 11) and in experiment 6 (n = 7) were homogenous: only infants at risk for DD (developmental dyslexia) were included. I will thus discuss the results of the two groups together. They main findings are summarized below.

During the RAP task, atypical infants showed the same auditory thresholds as their typical peers, but learned the operant conditioning procedure more slowly and less efficiently than the typical ones. They needed significantly more trials in order to reach the test phase.

Performing more trials during the association phase reflects a higher number of inattentive trials and generally poorer/slower discrimination abilities.

Nevertheless, our results do not converge with the previous literature, where infants at risk exhibited higher (thus poorer) auditory thresholds (e.g. Benasich & Tallal, 1996). The mean auditory thresholds measured here in infants at-risk were not significantly worse than those of their typical peers, most probably because typical infants' thresholds were underestimated. At the same time, the auditory thresholds of atypical infants were quite similar in an absolute value to those previously reported for infants at-risk (148 ms, in Benasich and Tallal, 1996). The lack of difference cannot be attributed to the small sample size as previous studies reported significant results testing similar numbers of atypical participants.

During the visual novelty detection task, the % of novelty effect, even if above chance, was significantly lower in the population at risk than in typically developing participants. As already discussed, this finding is in line with previous evidence proposing that this variable is predictive of later cognitive outcomes in both typical and atypical participants (e.g. Rose and Feldman, 1995; McCall and Carriger, 1993). This also converges well with the fact that atypical infants were slower to reach criterion in the association phase of the study, suggesting less efficient processing.

The mean vocabulary score in production but not in comprehension was significantly lower both at 14 months and at 18 months. Hence, measures of expressive vocabulary can be used longitudinally to evaluate the differential linguistic growth between typical and atypical populations. Importantly, the extent to which differences in early production are related to differences in early comprehension has to be better understood in future studies. For instance, some studies comparing 2-year-olds revealed that toddlers who showed a delay in production were not delayed in comprehension (Fischel et al., 1989). Hence, perhaps, delayed comprehension and production abilities may arise for at least partially different reasons, which should be investigated and interpreted more specifically.

Lastly, infants at-risk did not recognize infrequent words in an artificial grammar speech stream, while typical infants did. As the ability to make the "familiarity-to-novelty shift" depends on speed of processing (e.g. Hunter and Ames, 1988), infants in this group, perhaps, would have needed more repetitions or additional time in order to retrieve the familiar information. Again, this result converges with atypical infants slower and less efficient processing in the association phase and novelty detection of the RAP study.

In line with the existing literature, these findings seem to support the idea that some early behavioural markers of language risk/delays exist. The processing skills measured by these tasks may contribute to explain later language variability.

Succeeding in the psychoacoustic task requires the ability to discriminate, consolidate and retain in memory important information. Moreover, infants need to maintain stable attention over time. This might suggest the implication of short, long and working memory together (Choudhury et al., 2007).

By contrast, the habituation/novelty detection task requires discrimination of the familiar stimulus and the ability to disengage attention from a stimulus that is familiar to a stimulus that is unfamiliar and thus more informative (e.g. McCall, 1994; McCall & Mash, 1995, Sigman et al., 1997). Showing difficulties in the familiarity/novelty shift might reflect slower abilities to maintain and store short-term memory information (Choudhury et al., 2007).

In the case of the visual novelty detection task, even if not as efficient as typical peers, infants at-risk successfully habituated to the familiar stimulus and showed a novelty preference. Hence, they were able to use stored knowledge for a comparative purpose. In experiment 6, by contrast, they showed no preference. The null preference could arise from a weaker ability to encode the infrequent words during the familiarization phase, and/or to store them in memory, and/or to recall them during test or to compare them with the novel words in test. Perhaps, in both the visual novelty detection task and in experiment 6, atypical infants were able to discriminate the novel stimulus; nevertheless they still looked longer to the familiar. Weaker ability to disengage the attention from the familiar stimulus and/or slower ability to encode, store and retrieve relevant information might have been the possible causes of this different looking behaviour. Nevertheless, distinguishing whether infants had difficulty encode the words during the familiarization/habituation phase or they, instead, had problems recalling them during the test phase, is challenging, as looking time only offers a post hoc measure. Further investigation might involve the evaluation of these two specific skills separately. For instance, measures of attention abilities, short and long memory skills, and inhibitory control might be correlated with the early processing skills measured here.

Overall, this work provided relevant information about behavioural markers of DD that might be used in early screening.

The group differences that we found between the typical and atypical populations cannot be use as clinical, diagnostic markers of language delay. For the identification of a diagnostic marker, measures are needed that are accurate at the individual level. No overlap between the distributions for these markers in the typical and atypical populations should have occurred. Importantly, data from a larger number of participants at-risk need to be collected to compare more accurately the behavioural differences between the two populations.

Follow-up work thus needs to address whether these differences stay stable until the manifestation of language difficulties. Following these cohorts longitudinally might be informative in determining how many of these children end up with language impairments at school age. This might provide insight into the generalizability of predictive power of these risk factors.

4.3 Methodological strengths, limitations & follow-up studies

In the discussion of chapter 2, methodological limitations encountered during the Conditioned Head Turn Procedure (CHTP) were extensively discussed (section 5.4.1). The CHTP is a widely used methodology to assess individual speech perception abilities from infancy until childhood. This procedure not only allows measuring individual variability between performers, but it can also be used to track how these abilities change during development. In fact, the design is easily modifiable to assess differences and similarities in auditory/discriminative skills at different ages (see Werker et al., 1997 for a review). This procedure is also relevant to detect early difficulties in atypical populations (e.g. Benasich and Tallal, 1996).

Nevertheless, the CHTP, as many other experimental paradigms, also has some important disadvantages. First of all, it generally takes considerable time to complete the procedure. In the design applied here, each participant went through two learning phases (shaping and association) in order to enter the test phase where the acoustic threshold was measured. Hence, successful performance required discrimination of the two acoustic stimuli, but also the ability to learn the contingency between each stimulus and the correct video reward as well as the detection of temporal changes (decreased ISIs) in the acoustic sounds during the test phase. Thus, in addition to the acoustic processing, cognitive processes such as attention and generalization play a fundamental role in this task.

Within the tone and the syllable cohorts the 30.4% and the 42.8%, respectively, of the infants provided no data. The major causes were inability to reach the test phase and failure to reach the 4 reversal trials during the test phase. One possible, but costly, solution to reduce the rejection rate could have been to collect data at several time points from each infant at 9 months. Repeated sessions might allow each participant to learn first the procedure and then to be tested on the individual acoustic threshold.

To avoid the significant rejection rate, modified versions of this procedure have been proposed. Choudhury and colleagues (2007) have proposed a Go/ No-Go Operant Head Turn (HT) procedure (based on Morrongiello & Trehub, 1987; Trehub, Schneider & Henderson, 1995) to measure rapid auditory processing abilities. Infants were presented with the same tone stimuli that have been used here (100-100Hz; 100-300Hz) with ISIs at 500, 300 and 70 ms. The d-prime score, obtained by calculating the individual distribution between the hits and false alarms, was used as the detection index of the target tone at each ISI (from Choudhury et al., 2007). However, even if more baby-friendly, this modification does not use an adaptive procedure. Therefore, an informative portion of the individual variability is lost.

As language emerges from the interaction of many factors, this study empathises the importance of controlling for non-linguistic factors (e.g. general cognitive measures, etc.) when language abilities are assessed. The existence of correlations between general cognitive skills and later language outcomes implies that, by combining linguistic and cognitive measures, a more subtle understanding of individual language development may be obtained.

Moreover, to avoid the loss of meaningful individual differences, the standardized procedure for scoring language abilities using percentiles (CDIs) was not applied. Instead, raw scores were used. This procedure might represent a more sensitive approach to language outcomes across participants and ages. Nevertheless, in the syllable cohort, CDI measures were only partially reliable. Perhaps, different caregivers have filled out the questionnaire at different time points for the same infant, which might have lead to this unstable outcome. More precise results might be collected, for instance, by comparing the two parental evaluations and checking for correlations.

Following this idea, one way to extend the frequency-based bootstrapping study is to test whether performance on this task is linked to later language outcomes. As extensively reported in the general introduction, several studies in the literature have already demonstrated a link between early perceptual abilities, tested with the HPP paradigm, and later language outcome (e.g. Cristia & Seidl, 2011; Newmann et al., 2006). Measures of expressive vocabulary might be collected longitudinally at 36 months in the five groups of infants tested here. Children's first multiword production is telegraphic as functors are often omitted (Guasti, 2002). Nevertheless, as it was also confirmed here, the early omission of functors does not reflect a limitation on their perception or encoding (see also Gerken et al., 1990; Gerken & McIntosh, 1993). At 36 months, it might be interesting to investigate how early production of functors varies across learners and whether the magnitude of infants' preference for native word order is correlated with individual grammatical abilities. A positive correlation between these two variables might confirm the importance of functors in grammar learning and that efficiency in this discrimination contributes to later structural knowledge.

A second possible addition is to test participants of OV languages (e.g. Japanese, Turkish, Basque) in the experiments 4, 5 and 7. Young learners of OV languages should demonstrate an opposite word order preference switching from the frequent-initial to the frequent-final order. Moreover, they should also exhibit the same sensitivity to the open/closed lexical categorization. Hence, the preference for frequent-final order should be maintained when infrequent items are replaced and no preference is expected when novel items replace the frequent words.

Finally, in chapter 3, atypical infants only participated in experiment 6. The initial plan was to collect data from atypical participants in all the experiments. Nevertheless, time limitations and difficulties in recruiting a comparable number of infants at-risk made this impossible. Collecting comparative data from infants at risk for DD will provide important information regarding their sensitivity to word frequency as a cue to word order and the early distinction of lexical categories.

A limitation of the current work is that longitudinal studies provide correlational, but not causal information. Two variables can be correlated but not causally related, as they might both be dependent on a number of other variables that were not possible to measure.. Correlational results, in general, need to be interpreted carefully. Here, possible interpretations were proposed. However, additional data might be necessary to support these specific hypotheses.

4. Conclusion & perspectives

The present thesis contributed to the understanding of infants' speech perception abilities and how these abilities are related to later language outcomes. Even if far from being conclusive, several important results have been obtained. In line with the initial prediction, it proposed bootstrapping mechanisms as a mechanism that guides infants going from surface cues present in the input towards lexical and grammatical knowledge. Nevertheless, it also demonstrated that the rate and the efficiency at which infants process these surface cues exhibit great and informative individual variation, especially between typical and atypical populations.

Importantly, future research following infants longitudinally is needed to better understand the relations between early perception and later language development. This in turn might help to design more efficient paradigms that can be used for clinical purposes. At least four behavioural measures were proposed here as possible markers of language delay/impairments. Detecting early differences between typical and atypical populations is particularly useful to achieve more timely and efficient interventions.

5. Bibliography

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and language*, *162*, 46-59.

Abney, S. (1987). The English Noun Phrase in its Sentential Aspect. Unpublished PhD, MIT, Cambridge, MA.

Albareda-Castellot, B., Pons, F. & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental Science*, 14: 395-401

Aslin, R. N. (1989). Discrimination of frequency transitions by human infants. *The Journal of the Acoustical Society of America*, 86 (2): 582-90.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10 (1), 48–53.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. Handbook of child psychology: formerly Carmichael's Manual of child psychology/Paul H. Mussen, editor.

Bateson, P. (1979). How do sensitive periods arise and what are they for?. *Animal Behaviour*, *27*, 470-486.

Bavelier, D., Davidson, R. J. (2013). Brain training: Games to do you good. *Nature,* 494:425–26.

Bayley, N. (1949). Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years. *The Pedagogical Seminary and Journal of Genetic Psychology*, *75*(2): 165–196.

Bayley, N. (1969). Manual for the Bayley scales of infant development. Psychological Corporation.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255-309.

Benasich, A. A. & Tallal, P. (1998). Infant processing of auditory temporal information: links to family history and later language outcome. *Society for Neuroscience*, 24:819.

Benasich, A. A., & Leevers, H. J. (2002). Processing of rapidly presented auditory cues in infancy: Implications for later language development. In J. Fagan & H. Haynes (Eds.), *Progress in infancy research: Vol. 3* (pp. 245–288). Mahwah, NJ: Lawrence Erlbaum Associates.

Benasich, A. A., & Tallal, P. (1996). Auditory temporal processing thresholds, habituation, and recognition memory over the 1st year. *Infant Behavior and Development*, 19, 339–357.

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural Brain Research*, 136 (1), 31–49.

Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., & Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia,* 44 (3): 396-411.

Berg, C. A., & Sternberg, R. J. (1985). Response to Novelty: Continuity Versus Discontinuity in the Developmental Course of Intelligence. In H. W. Reese (Ed.), 19, 1–47.

Berg, K. M., & Smith, M. C. (1983). Behavioral thresholds for tones during infancy. *Journal of Experimental Child Psychology*, 35 (3), 409-425.

Bernard, C. & Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Frontiers in Psychology*, 3, 451.

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: distinguishing transient from persistent impair- ment. *Journal of Speech and Hearing Disorders*, 52, 156–173.

Bishop-Liebler, P., Welch, G., Huss, M., Thomson, J. M., & Goswami, U. (2014). Auditory temporal processing skills in musicians with dyslexia. *Dyslexia*, *20*(3), 261-279.

Bornstein, M. H. & Sigman, MD. (1986). Continuity in mental development from infancy. *Child Development*, 57, 251- 274.

Braine, M. D. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, 72 (4), 532-540.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review,* 70 (4), 323-348

Brown, C. M., Hagoort, P. & Kerus, M. (1999). Electrophysiological Signatures of Visual Lexical Processing: Open-and Closed-Class Words. *Journal of Cognitive Neuroscience*, 11 (3), 261-281.

Brown, R. W. (1973). A First Language; the Early Stages. Cambridge, MA: Harvard University Press.

Bruner, J. (1983) Child's talk: Learning to use language. New York: W.W. Norton

Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition,* 63(2), 121-70.

Chang, H. W., & Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child development*, 1666-1670.

Chomsky, N. (1959). Review of Skinner's Verbal Behavior. *Language*, 35(1), 26–58

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, 3(1), 1-15.

Chomsky, N. (1981). Principles and parameters in syntactic theory. *Explanation in linguistics*, 32-75.

Chomsky, N. (1986). *Barriers* (Vol. 13). MIT press.

Chomsky, N. (1995). The Minimalist Program. Cambridge: MIT Press.

Choudhury, N., Leppanen, P. H., Leevers, H. J., & Benasich, A. A. (2007). Infant information processing and family history of specific language impairment: converging evidence for RAP deficits from two paradigms. *Developmental science*, *10*(2), 213-236.

Christophe, A., Guasti, T., & Nespor, M. (1997). Reflections on Phonological Bootstrapping: Its Role for Lexical and Syntactic Acquisition. *Language and Cognitive Processes*, 12 (5-6), 585–612.

Christophe, A., Nespor, M., Guasti, T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, 6, 211–220.

Colombo, J., Mitchell, D. W., Dadd, J., Coldren, J. T., & Horowitz, F. D. (1989). Longitudinal correlates of infants attention in the paired-comparison paradigm. *Intelligence*. 13, 33- 42.

Cristia, A., & Seidl, A. (2011). Sensitivity to prosody at 6 months predicts vocabulary at 24 months. In N. Denis, K. Mesh, & H. Sung (Eds.) Proceedings of the 35th BUCLD (pp. 145–156). Somerville, MA: Cascadilla Press.

Cutler, A. (1993). Phonological cues to open-and-closed-class words in the processing of spoken sentences. *Journal of Psycholinguisitic Research*, 22 (2).

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, 40 (2), 141-201.

Dale, P. S, Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T. S., Purcell, S., Stevenson, J., Plomin, R. (1998) Genetic influence on language delay in 2-year-old children. *Nature Neuroscience*, 1(4): 324.

Davis, S. M., & McCroskey, R. L. (1980). Auditory fusion in children. *Child Development,* 5 (1), 75-80.

de la Cruz-Pavía, I., Elordieta, G., Sebastián-Gallés, N. & Laka, I. (2015). On the role of frequency-based cues in the segmentation strategies of adult OV- VO bilinguals. *International Journal of Bilingual Education and Bilingualism*, 18 (2), 225-241

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Spoken Language, 1996. ICSLP 96. Proceedings, Fourth International Conference on* 3,1393-1396. IEEE.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental Changes in Speech Discrimination in Infants. Journal of Speech and Hearing Research, 20(4), 766-780.

Elliot, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America,* 66, 651-653.

Elliott, L. L & Katz, D. R. (1980). Children's pure tone-detection. *The Journal of the Acoustical Society of America,* 67 (1), 343-344.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, 2, 460–493.

Fenson, L. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). Technical manual for the MacArthur communicative development inventory. *San Diego: Singular Press.*

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, 42(1), 98.

Fior, R. (1972). Physiological Maturation of Auditory Function Between 3 and 13 Years of Age. *Audiology,* 11 (5), 317-321.

Fischel, J. E., Whitehurst, G. J., Caulfield, M. B., & DeBaryshe, B. (1989). Language growth in children with expressive language delay. *Pediatrics*, *83*(2), 218-227.

Fitch, R. H., Read, H., & Benasich, A. A. (2001). Neurophysiology of speech perception in normal and impaired systems. *Physiology of the ear*, 651-672.

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (in press, 2018). Vocabulary of 2-year-olds learning English and an Additional Language: Norms and effects of linguistic distance. Monographs of the Society for Research in Child Development, Serial No. 328, Vol. 83, No. 1.

Foxe, J. J., & Stapells, D. R. (1993). Normal infant and adult auditory brainstem responses to bone-conducted tones. *Audiology*, 32(2), 95-109.

Friedrich, M., Herold, B., & Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. Cortex, 45, 662–676.

Friedrich, M., Weber, C., & Friederici, A. D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*, *41*, 772–782.

Fukui, N. (1986). A Theory of Category Projection and its Applications. Unpublished PhD, MIT, Cambridge, MA.

Galaburda, A. M., & Kemper, T. L. (1979). Cytoarchitectonic abnormalities in developmental dyslexia: a case study. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 6(2), 94-100.

Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitiz, F., & Geschwind, N. (1985). Developmental dyslexia: four consecutive patients with cortical anomalies. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *18*(2), 222-233.

Gerken, L. A & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29(3), 448-457.

Gerken, L. A., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, 26, 204–216.

Gerrits, E., & de Bree, E. (2009). Early language development of children at familial risk of dyslexia: Speech perception and production. *Journal of Communication Disorders*, *42*(3), 180-194.

Gervain, J. (2007). "From the Richness of the Signal to the Poverty of the Stimulus: Early Mechanisms of Language Acquisition" PhD diss., SISSA.

Gervain, J. & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications,* 4, 1490.

Gervain, J., Nespor, M., Mazuka, R., Horie, R. & Mehler, J. (2008). "Bootstrapping Word Order in Prelexical Infants: A Japanese-Italian Cross-linguistic, Study." *Cognitive Psychology*, 57, 56–74.

Gervain, J., Sebastián-Gálles, N., Díaz, B., Laka, I., Mazuka, R., Yamane, N., Nespor, M. & Mehler, J. (2013). Word Frequency Cues Word Order in Adults: Cross- linguistic Evidence. *Frontiers in Psychology*, 4, 689.

Gilger, J. W., Pennington, B. F., & Defries, J. C. (1991). Risk for reading disability as a function of parental history in three family studies. In *Reading disabilities* (pp.17-29). Springer, Dordrecht.

Gleitman, L. (1990) The structural sources of verb meaning. *Language Acquisition*, 1, 3-55.

Gleitman, L. R., & Wanner, E. (1982). *Language acquisition: The state of the art*. CUP Archive.

Goldberg, A. E. (1995). Constructions: A costruction grammar approach to argument structure. University of Chicago Press.

Goldberg, A. E. (2006). Constructions at work: The nature of generalization in language. Oxford University Press on Demand.

Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental science*, *15*(6), 885-894.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives. *Lab. Phonol*, *4*(1), 67-92.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. Reading and Writing, 23, 995–1019.

Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn and infant nursing reviews*, 8(4), 187-193.

Green, T. R. G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, 18, 481- 496.

Griffiths, S. K., Brown Jr, W. S., Gerhardt, K. J., Abrams, R. M., & Morris, R. J. (1994). The perception of speech sounds recorded within the uterus of a pregnant sheep. *The Journal of the Acoustical Society of America*, 96(4), 2055-2063.

Grimm, H., Aktas, M., & Frevert, S. (2001). SETK 3–5: Sprachentwicklungstest für drei-bis fünfjährige Kinder [Language development test for three to five year old children; German version]. Göttingen: Hogrefe.

Guasti, M. T. (2002). Language Acquisition: The Growth of Grammar. Cambridge, MA: MIT Press.

Guttorm, T. K., Leppänen, P. H. T., Poikkeus, A. M., Eklund, K. M., Lyytinen, P., & Lyytinen, H. (2005). Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex*, 41(3): 291–303.

Guttorm, T. K., Leppänen, P. H. T., Richardson, U., & Lyytinen, H. (2001). Event-related potentials and consonant differentiation in new- borns with familial risk for dyslexia. *Journal of Learning Disabilities*, 34: 534–544.

Guttorm, T. K., Leppänen, P. H. T., Tolvanen, A., & Lyytinen, H. (2003). Event-related potentials in newborns with and without familial risk for dyslexia: Principal component analysis reveals differences between the groups. *Journal of Neural Transmission*, *110*(9), 1059-1074.

Hämäläinen J. A., Salminen H. K., & Leppänen P. H. T. (2012). Basic auditory processing deficits in dyslexia: Systematic review of the behavioural and event-related potential/ field evidence. Journal of Learning Disabilities.

Hayes, B. (1995). Metrical stress theory: Principles and case studies. Chicago, IL: University of Chicago Press.

Hayes, B. (1989). The prosodic hierarchy in meter. In *Rhythm and meter* (pp. 201-260).

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience* 6:877–88.

Hepper, P. (1991). An examination of fetal learning before and after birth. *The Irish Journal of Psychology*, **12**, 95–107.

Hepper, P. G., & Shahidullah, B. S. (1994). The development of fetal hearing. *Fetal and Maternal Medicine Review*, *6*(3), 167-179.

Hochmann, J. R., Endress, A. D. & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115, 444–457

Höhle, B., Pauen, S., Hesse, V., & Weissenborn, J. (2014). Discrimination of Rhythmic Pattern at 4 Months and Language Performance at 5 Years: A Longitudinal Analysis of Data From German-Learning Children. *Language Learning*, *64*(2), 141-164.

Höhle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, 2, 277–300.

Horowitz, F. D., Paden, L., Bhana, K. & Self, P. (1972). An infant-controlled procedure for studying infant visual fixations. *Developmental Psychology*, 7, 90.

Houtgast, T., & Steeneken, H. J. M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica*, 28, 66-73.

Humphreys, P., Kaufmann, W. E., & Galaburda, A. M. (1990). Developmental dyslexia in women: neuropathological findings in three patients. *Annals of Neurology: Official*

Journal of the American Neurological Association and the Child Neurology Society, 28(6), 727-738.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine & Child Neurology*, *32*(4), 352-355.

Hutt, S. J.,Hutt, C., Lenard, H. G., Bernuth, H., Muntjewerff, W. J. (1968) Auditory responsiveness in the human neonate. Nature, 218, 888-90.

Irwin, R. J., Ball, A. K. R., Kay, N., Stillman, J. A., & Rosser, J. (1985). The development of auditory temporal acuity in children. *Child Development,* 56, 614-620.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Johnston, R., Stark, R., Mellits, E., & Tallal, P. (1981). Neurological status of language-impaired and normal children. *Annals of Neurology*, 10, 159–163.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A., & Myers, M. (1983). Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science*, *222*(4620), 175-177.

Jusczyk, P. W., Pisoni, D. B., Walley, A., Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *The Journal of the Acoustical Society of America*, 67: 262-270.

Kandel, E. R., Schwartz, J. H., Jessel, T. M., Siegelbaum, S. A., Hudspeth, A. J. (2012). Principles of Neural Science, Fifth Edition. McGraw Hill Professional.

Kopp, C. B., McCall, R. B. (1982). Predicting later mental performance for normal, at-risk and handicapped infants. In P. B., Bates, O. G Brim, (Eds), *Life-span development and behaviour*, 4, 33-61.

Kucera, H. & Francis, W. N. (1967). Computational Analysis of Present-day American English. Providence: Brown University Press.

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron* 67, 713–27.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pru- itt, J. (2005). Early speech perception and later language development. Language Learning and Development, 1, 237–264.

Lenard, H., Von Bernuth, H., & Hutt, S. J. (1969). Acoustic evoked responses in newborn infants: the influence of pitch and complexity of the stimulus. *Clinical Neurophysiology*, *27*(2), 121-127.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59-67.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Weismer, S. E., Miller, C., Francis, D. J., Tomblin, J. B., Kail, R.V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*; 50, 408–428.

Leppänen, P. H., Richardson, U., Pihko, E., Eklund, K. M., Guttorm, T. K., Aro, M., & Lyytinen, H. (2002). Brain responses to changes in speech sound durations differ between infants with and without familial risk for dyslexia. *Developmental neuropsychology*, *22*(1), 407-422.

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49 (2), 467–477.

Lightfoot, D. (1991). How to set parameters: Arguments from language change. Cambridge, MA, US: The MIT Press.

Litovsky, R. (2015). Development of the auditory system. In *Handbook of clinical neurology, 129*, 55-72 Elsevier.

Maratsos, M. P. & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K.E. Nelson (Ed.), Children's language, Gardner Press, New York, 2, 127-214

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. Developmental Psychology, 28(3), 396.

Marquis, A., & Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, 122, 61–66.

McCall, R. B. (1979). The development of intellectual functioning in infancy and the prediction of later IQ. In J. Osofky (Ed.) *Handbook of infants development.* New York: Wiley, 707-741.

McCall, R. B. (1994). What process mediates predictions of childhood IQ from infant habituation and recognition memory? Speculations on the roles of inhibition and rate of information processing. *Intelligence*, 18 (2), 107–125.

McCall, R. B., & Carriger, M.S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, 64, 57–79.

McCall, R. B., & Mash, C.W. (1995). Infant cognition and its relation to mature intelligence. In R. Vasta (Ed.), *Annals of child development: A research annual* (Vol. 10, pp. 27–56). London: Jessica Kingsley.

Michel, G. F., & Moore, C. L. (1995). Developmental psychobiology: An interdisciplinary science. MIT Press.

Moore, D. R., Ferguson, M. A., Halliday, L. F. & Riley, A. (2008). Frequency discrimination in children: perception, learning and attention. *Hearing Research* 238, 147–154.

Morrongiello, B.A., & Trehub, S.E. (1987). Age-related changes in auditory temporal perception. *Journal of Experimental Child Psychology*, 44, 413–426.

Morgan, J. L. & Demuth, K. (1996). Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 263–283.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 263-283.

Morgan, J. M. & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 20 (1): 67-85

Morgan, J. M., Meier, R. P. & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19 (4), 498-550.

Mullen, E. M. (1995). Mullen scales of early learning. Circle Pines, MN: AGS, 58-64.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779-784.

Nelson, D. G., Jusczyk, P., Mandel, D., Myers, J., Turk, A. & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant behaviour and development*, 18 (1), 111-116.

Nelson, K. (1996) Language in cognitive development. New York: Cambridge University Press.

Nespor, M. & Vogel, I. (2007). Prosodic phonology 2nd edition.

Nespor, M. & Vogel, I. (1986). Prosodic Phonology, Vol. 28 Dordrecht: Foris, 1st edition.

Nespor, M., Guasti, M.T., & Christophe, A. (1996) Selecting word order: The rhythmic activation principle. In U. Kleinhenz, Interface in Phonology. Berlin: Academic Verlag

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H. & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 7, 1–28.

Neuman, A. C., & Hochberg, I. (1983). Children's perception of speech in reverberation. Journal of the Acoustical Society of America, 73, 2145-2149.

Newman, R. S., Bernstein Ratner, N., Jusczyk, A. M., Jus- czyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development. Developmental Psychology, *42*, 337–367.

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of communication disorders*, *32*(4), 223-245.

Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., & Spetner, N. B. (1988). Pure-tone sensitivity of human infants. The Journal of the Acoustical Society of America, 84(4), 1316-1324.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385-413.

Pennington, B. F., Gilger, J. W., Pauls, D., Smith, S. A., Smith, S. D., & DeFries, J. C. (1991). Evidence for major gene transmission of developmental dyslexia. *Jama*, *266*(11), 1527-1534.

Piaget, J. (1952) The origins of intelligence in children (New York, International Universities Press).

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, 271-311.

Pinker, S. (1984-1996). Language learnability and language development. Cambridge, MA: Harvard.

Plomin, R., DeFries, J. C., McClearn, G. E. (1990) Behavioral genetics: a primer, 2nd ed.. New York: Cambridge University Press.

Pujol, R., & Lavigne-Rebillard, M. (1992). Development of neurosensory structures in the human cochlea. *Acta oto-laryngologica*, *112*(2), 259-264.

Radford, A. (1990) Syntactic theory and the acquisition of English syntax: the nature of early child grammars of English*. Oxford: Basil Blackwell.*

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6 (6), 855-863.

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of experimental child psychology*, *48*(2), 270-292.

Richards, S. & Goswami, U. (2015). Auditory processing in Specific Language Impairment (SLI): Relations with the perception of lexical and phrasal stress. *Journal of Speech and Language Hearing Research;* 58(4), 1292-305.

Richardson, U., Leppänen, P. H., Leiwo, M., & Lyytinen, H. (2003). Speech perception of infants with high familial risk for dyslexia differ at the age of 6 months. *Developmental neuropsychology*, *23*(3), 385-397.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. NeuroReport (Developmental Neuroscience), 16, 495–498.

Rizzi, L. (1986). Null Objects in Italian and the Theory of *pro*. *Linguistic Inquiry* 17 (3), 501-557.

Roche, A. F., Siervogel, R. M., Hines, J. H & Johnson, D. L. (1978). Longitudinal study of hearing in children: Baseline data concerning auditory thresholds, noise exposure, and biological factors. *Journal of the Acoustical Society of America*, 64, 1593-1601.

Rose, S. A., & Feldman, J. F. (1990). Infant cognition: Individual differences and developmental continuities. In J. Colombo & J. W. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*, 229-245. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Rose, S. A., & Feldman, J. F. (1995). Prediction of IQ and specific cognitive abilities at 11 years from infancy measures. *Developmental Psychology*, *31*(4), 685.

Rose, S. A., Feldman J. F. & Jankowski J. J. (2009). A cognitive approach to the development of early language. *Child Development*, 80 (1), 134-150.

Rose, S. A., Feldman, J. F. & Jankowski, J. J. (2005a). Recall memory in the first three years of life: A longitudinal study of preterms and full-terms. *Developmental Medicine and Child Neurology*, 47, 653–659.

Saffran, J. R., Johnson, E. K., Aslin, R. N. & Newport, E. L. (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1), 27-52.

Scott, J. P. (1962). Critical periods in behavioral development. *Science*, *138*(3544), 949-958.

Scott, J. P., & Marston, M. V. (1950). Critical periods affecting the development of normal and mal-adjustive social behavior of puppies. *The Pedagogical Seminary and Journal of Genetic Psychology*, *77*(1), 25-60.

Selkirk, E. (1984) Phonology and syntax: The relation between sound and structure. Cambridge, MA: The MIT Press.

Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language and Hearing Research*, 41, 874–886.

Shi, R. & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11, 407–413.

Shi, R. & Melancon, A. (2010). Syntactic categorization in French- learning infants. *Infancy*, 15, 517–533.

Shi, R. & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, 12, 70–75.

Shi, R., & Werker, J. F. (2003). The basis of preference for lexical words in six-month-old infants. *Developmental Science*, 6, 484–488.

Shi, R., Cutler, A., Werker, J., Cruickshank, M. (2006a). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of Acoustic Society of America,* 11, 9(6).

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, 25, 169–201.

Shi, R., Werker, J. F. & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72.

Shi, R., Werker, J., & Cutler, A. (2006b). Recognition and representation of function words in English-learning infants. *Infancy*, 10, 187–198.

Shipley E. F., Smith C. S. & Gleitman L. R. (1969). A Study in the Acquisition of Language: Free Responses to Commands. *Language* 45 (2), 322-342.

Sigman, M., Cohen, S.E., & Beckwith, L. (1997). Why does infant attention predict adolescent intelligence? *Infant Behavior and Development*, 20, 133–140.

Singh, L., Reznick, J. R., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development. Developmental Science, 15, 482–495.

Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *The Journal of the Acoustical Society of America*, *78*(6), 1986-1992.

Skinner, B. (1957). *Verbal Behavior* (Appleton FC.). New York.

Smith, N. A., Trainor, L. J., & Shore, D. I. (2006). The development of temporal resolution: between-channel gap detection in infants and adults. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1104-1113.

Snowling, M. J. (2000). Dyslexia (2nd ed.). Oxford, England: Blackwell.

Stellmack, M. A., Willihnganz, M. S., Wightman, F. L. & Lutfi, R. A. (1997). Spectral weights in level discrimination by preschool children: analytic listening conditions. *Journal of Acoustic Society of America*, 101 (5), 2811–2821.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence.

Stoodley, C.J., Talcott, J.B., Carter, E.L., Witton, C., & Stein, J.F. (2000). Selective deficits of vibrotactile sensitivity in dyslexic readers. *Neuroscience Letters*, 295, 13–16.

Sutcliffe, P., & Bishop, D. (2005). Psychophysical design influences frequency discrimination performance in young children. *Journal of experimental child psychology*, *91*(3), 249-270.

Suzuki, T. & Ogiba, Y. (1971). Conditioned Orientation Reflex Audiometry: A New Technique for Pure-Tone Audiometry in Young Children Under 3 Years of Age. *Archives Otolaryngology*, 74 (2), 192–198.

Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics*, 5, 167–169.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Review in Neuroscience* 5, 721–728.

Tallal, P., & Curtiss, S. (1990). Neurological basis of developmental language disorders. In A. Rothenberger (Ed.), *Brain and behavior in child psychiatry*. New York: Springer-Verlag, 205–216.

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia*, 11: 389–398.

Tallal, P., Stark, R. E., & Mellits, D. (1985a). The relationship between auditory temporal analysis and receptive language development: Evidence from studies of developmental language disorder. *Neuropsychologia*, *23*(4), 527-534.

Tallal, P., Stark, R. E., & Mellits, E. D. (1985b). Identification of language-impaired children on the basis of rapid perception and production skills. *Brain and language*, *25*(2), 314-322.

Tellegen, P.,Winkel, M.,Wijnberg-Williams, B., & Laros, J. (1998). Snijders-Oomen Non-verbaler Intelligenztest SON-R 2½ –7.Manual. Lisse, Netherlands: Swets & Zeitlinger B. V.; Swets Test Services Frankfurt.

Thompson, L., Fagan, J., & Fulker, D. (1991). Longitudinal Prediction of Specific Cognitive Abilities from Infant Novelty Preference. *Child Development*, 62(3), 530-538.

Thorndike, R. L., Hagen, E. P., Sattler, J. M. (1986). The Stanford binet intelligence scale, 4th ed. Chicago, IL: The Riverside Publishing Company.

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology*, *25*(1), 122.

Trehub, S.E., Schneider, B.A., & Henderson, J.L. (1995). Gap detection in infants, children, and adults. *Journal of the Acoustical Society of America*, 98, 2532–2541.

Tomasello, M. (1992) The social bases of language acquisition. *Social Development* 1(1), 67-87.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209–253.

Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press.

Toro J. M., Nespor M. & Gervain J. (2016). Frequency-based organization of speech sequences in a nonhuman animal. *Cognition,* 146, 1-7.

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception & psychophysics*, *62*(2), 333-340.

Trehub, S. E., & Henderson, J. L. (1996). Temporal resolution in infancy and subsequent language development. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1315-1320.

Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *The Journal of the Acoustical Society of America*, *98*(5), 2532-2541.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child development*, *75*(4), 1067-1084.

Valian, V. & Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory and Language*, 27, 71-86.

Vygotsky,L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Waxman, S. R., Lidz, J., Braun, I. E., & Lavin, T. (2009). Twenty-four- month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, 59, 67–95.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiologial evidence. *Cognitive Brain Research*, *25*(1), 180-187.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annual review of psychology*, 66.

Werner, L. A., & Gillenwater, J. M. (1990). Pure-tone sensitivity of 2-to 5-week-old infants. *Infant Behavior and Development*, *13*(3), 355-375.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, *24*(5), 672.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *37*(2), 278.

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., & Gillenwater, J. M. (1992). Infant auditory temporal acuity: Gap detection. *Child Development*, 63; 260–272.

Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development*, 611-624.

Zimmerman, I. L, Steiner, V. G, Pond, R. E. (1992). Presschool language scale-3. New York: The Psychological Corporation.

Zipf, G. (1953). The psycho-biology of language. Boston, MA: Houghton Mifflin.

6. Annexes

Non-significant correlations discussed in chapter 2 are presented in the following tables. Due to the huge amount of data and variables, only theoretically relevant comparisons are presented.

6.1 Pearson correlations for auditory temporal threshold and habituation/visual novelty detection variables (tone cohort).

Exp. 1: Tones	% novelty effect	TTC	First looking length (s)	% Response Decrement	Total looking time (s)	Habituation Slope (α)
Four best succeeded	0.212	-0.168	-0.346	-0.234	-0.249	-0.042
Four last reversals	0.232	-0.180	-0.183	-0.104	-0.206	-0.043
Four best reversals	0.276	-0.170	-0.128	-0.195	-0.149	0.056
n trials association	-0.067	-0.092	0.202	0.029	0.114	0.224
n trials test phase	-0.087	-0.325	0.020	0.097	-0.253	0.191
Total n of trials	-0.098	-0.292	0.107	0.088	-0.143	0.249

6.2 Pearson correlations for auditory temporal threshold and habituation/visual novelty detection variables (syllable cohort)

Exp. 2: Syllables	% novelty effect	TTC	1^{st} looking length	% Decrement Response	Total looking time	Slope α
Four best succeeded	-0.085	-0.106	0.050	-0.017	-0.127	0.0222
Four last reversals	-0.080	-0.150	-0.002	-0.182	-0.143	0.005
Four best reversals	-0.068	-0.108	-0.0262	-0.201	-0.115	-0.005
n trials association	0.097	0.220	0.054	0.264	0.103	0.063
n trials test phase	0.189	0.123	0.148	0.431*	0.2846	0.206
Total n of trials	0.161	0.179	0.124	0.396	0.229	0.162

6.3 Pearson correlations for auditory temporal threshold variables and raw scores in comprehension and production (tone and syllable cohorts)

Tone cohort	Score comprehension 9 months	Score comprehension 12 months	Score comprehension 14 months	Score production 18 months	Score comprehension 24 months	Score production 24 months
Four best succeeded	-0.093	0.01	0.204	0.215	0.095	0.113
Four last reversals	-0.122	0.077	0.272	0.227	0.107	0.07
Four best reversals	-0.061	0.106	0. 323	0.226	0.067	0.029

Syllable cohort	Score comprehension 9 months	Score comprehension 12 months	Score comprehension 14 months	Score comprehension 18 months	Score production 18 months
Four best succeeded	-0.081	0.021	0.036	0.317	0.182
Four last reversals	-0.012	0.027	0.124	0.270	0.102
Four best reversals	-0.051	-0.023	0.118	0.218	0.126

6.4 Pearson correlations for habitation/visual novelty detection variables and raw scores in comprehension and production at 18 months (tone cohort).

Tone cohort 18 months	TTC	First looking length (s)	% response decrement	Total looking time (s)	Habituation slope (s)
Comprehension score	0.302	0.307	0.209	0.194	-0.268
Production score	0.141	0.364	-0.02	-0.255	-0.671***

6.5 Pearson correlations for habitation/visual novelty detection variables and raw scores in comprehension and production at 18 months (syllable cohort).

Syllable cohort 18 months	TTC	First looking length (s)	% response decrement	Total looking time (s)	Habituation slope (s)
Comprehension score	-0.127	-0.120	0.246	-0.070	-0.144
Production score	-0.169	0.079	0.163	-0.072	0.094