Scuola Internazionale Superiore di Studi Avanzati
International School for Advanced Studies

Word learning in the first year of life

Candidate:
Amanda Saksida

Supervisors:
Jacques Mehler
Marina Nespor
Alan Langus

Thesis submitted for the degree of Doctor of Philosophy in Cognitive Neuroscience
Trieste, 2014

Committee members

Gergely Csibra
Cognitive Development Center, Department of Cognitive Science,
Central European University, Budapest, Hungary

Luca Bonatti
Center for Brain and Cognition, Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Alessandro Treves
Liminar Investigations of Memory and Brain Organization Lab, Neuroscience Area
International School for Advanced Studies, Trieste, Italy

Raffaella Rumiati
Social Cognition & Integrative Neuroscience Lab, Neuroscience Area
International School for Advanced Studies, Trieste, Italy

Giosuè Baggio
Brain and Language Laboratory, Neuroscience Area
International School for Advanced Studies, Trieste, Italy

Acknowledgments

I would like to dedicate this thesis to my father, Iztok Saksida. He gave me the stubbornness, the self-irony, and the ludicrousness needed to do the big things in life correctly, although in a wrong order.

Needless to say life is an incredible journey, but these past four years indeed were an incredible journey. After years of teaching linguistics, translating, editing, and having a family, I knew it is time to start doing what I really wanted, studying language acquisition. So I applied for the study at SISSA, and did the exams. I thought it must have been some kind of a joke when I learned I'm accepted, but I liked the joke, and so I started to study here. I came to the lab without much research experience, without much knowledge about experimental psychology and even less about cognitive neurosciences. But I came to the best place possible, and learned more than I ever thought I would.

This would not be possible without Jacques who kindly accepted me in his lab in spite of all the existing counter-arguments, who took time to talk about great ideas and great people, and reminded me so many times that the devil is in the details. It would not be possible without Marina, who followed patiently my work with Alan Langus, and who boosted my efficiency. It would be totally impossible without Alan, who was, in many aspects, my supervisor too. We elaborated most of the ideas written in this thesis together. We spent numerous hours debating, analyzing, and writing together, and his creativity opened new horizons for me. I owe the three of them my deepest gratitude.

There are two students of this lab that will remain in my heart forever. Silvia Benavides Varela, who is probably one of the most kind and graceful human beings on earth, took time to teach me about experimental techniques, designs, and analysis any time I needed. There are not many people like her, and I'm honored to have met her. David Maximiliano Gomez, who has the rare talent to present clearly the most complex concepts, taught me everything I know about stats and data analysis. I thank them from for everything they have done for me.

But in fact, the whole community of alumni of this lab is special because it accepts new members with a sense of familiarity, which I appreciated a lot. Whenever I saw them, Jean-Remy Hochmann (who was on the leave when I came to SISSA), Agnes Kovacs, Erno Teglas, Judit Gervain, Juan Manuel Toro and Mohinish Shukla were always prepared to talk and share experiences, and I am grateful for that. In the last two years, the PASCAL project brought many new people to this lab. I am grateful to Ana Flo, Yamil Vidal Dos Santos, Milad Ekramnia, Alissa Ferry, Bahia Guellai, and Hanna Marno for the numerous small supports and discussions we had. This lab would function very differently without the two important and beautiful persons who made sure that our administrative issues were covered, that the testing pursued smoothly, and were always there with a warm word. Marijana Sjekloća and Francesca Gandolfo, I will never forget you! My life at SISSA would not have been the same without Alessio Isaja and Andrea Sciarrone, the administrative and informational life-saviors and my dear tango friends, and Ritwik Kulkarni, my friend and companion in the evenings spent in the office working. He got chickenpox from our kids, and our kids became vegetarian because of him – that’s life too :). Thank you all!

No journey is without the twists, and neither was this. The back injury in 2012 has, ironically, given me the opportunity to start learning to program. I will never forget the exciting hours I spent discovering the joys of programming, which would be much less joyful and more time consuming without my friend and teacher, Sašo Živanović. Thank you, Sašo! And when I came back to work, I started to share the office with an inspiring person, John Nicholls. He has challenged my views about nearly everything, has told numerous stories, and became my role model for ageing. I am grateful I had the honor to have met him.

Last, but definitely not least, this journey would be quite different without my family. Jona, Filip, and Loti have always reminded me that life is even much more complex than any developmental literature would suggest, and have given me the ground, the direction and the balance to always know what is really important. What is really important is the three of them, and I am grateful for their existence. And finally, the person who has loved me, challenged me, irritated me, stimulated me, supported me, loved me more, and has been the most important person in my life, is Pavel Koltaj. I couldn’t have started this journey without him, and I couldn’t have completed it without him. So a simple thank you is just not enough.

Znanost naj svet spreminja, ne samo razlaga. Prav zaradi tega se mi zdi ukvarjanje z znanostjo produktivno in zabavno. V znanost je mogoče uživati …

Iztok Saksida

(*Na poti v Socionomijo*, Ljubljana: Studia Humanitatis, 1999)

... ho imparato che niente è impossibile, e anche che quasi niente è facile ...

Un Urlo, Articolo 31

Table of contents

List of figures

List of tables

1

Introduction

Learning words is one of the most obvious tasks people face when learning a new language. When people think about their proficiency in a language, the first thing that comes to their mind is often the size of their vocabulary. To assess the level of knowledge in a foreign language, it is most common to test the size of one's vocabulary. Similarly, an important part of the surveys of first language development in infants and toddlers is constituted by the estimations of the size of infants' or toddlers' receptive and/or active vocabulary (Fenson et al., 1993). Therefore, the question is when infants start to learn words, and how. Some studies have indicated that infants may learn words well before their first birthday (Pruden, Hirsh-pasek, Golinkoff, & Hennon, 2006; Shukla, White, & Aslin, 2011). Recently, a study was conducted that shows that infants' receptive vocabulary in the first year of life is already quite impressive (Bergelson & Swingley, 2012) and that it includes many common nouns. Infants must learn all these words somehow. The environment in which they must learn words is a natural environment with numerous possible ambiguities. The words are, each time they are pronounced, characterized by at least slightly different physical properties. The items that are labeled also change location and the angle of view. Both visual and auditory input therefore must be categorized. Moreover, infants must extract words from fluent speech, map the correct word to the correct referent among many possible ones, and remember it. Research on word learning should study how young language learners achieve each of these tasks.

The literature on infant word learning is abundant and can be divided into four broad areas that correspond to the infant's tasks mentioned above: (1) the development of native language segmental categorization (Kuhl, 1991, 2004; Mehler, 1981; Ramon-Casas, Swingley, Sebastián-Gallés, & Bosch, 2009; Sebastián-Gallés, 2006); (2) the development of categorization of the visual input and its possible interaction with language (Ferry, Hespos, & Waxman, 2010, 2013; Hespos & Spelke, 2004; Quinn, Eimas, & Rosenkrantz, 1993); (3) segmentation of words from fluent speech (Aslin, Saffran, & Newport, 1998; Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996; Shukla, Nespor, & Mehler, 2007); (4) acquiring the meaning of words – conceptual mapping between the word form and the concept for which it stands (Carey, 2010; Golinkoff & Hirsh-Pasek, 2008; Shukla, White, & Aslin, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013, among others). In each of these areas of research researchers have formulated hypotheses about possible learning mechanisms involved in the different tasks. One of the main issues investigated concerns the extent to which these abilities are driven by innate capacities.

Some approaches follow the view that learning in humans is driven by sensory-motoric experiences, which are a necessary base for higher-level abstractions. They emphasized the role of perceptual saliency also in word learning tasks (Gibson, Owsley, & Johnston, 1978; Piaget & Cook, 1952). Just as the visual and auditory categorizations are driven by perceptual similarity, words are learned through infants' perception of invariant features in salient events from different perceptual domains (the salient auditory word and the salient visual event). Temporal audio-visual synchrony of the events in different modalities, which match in tempo and intensity, creates the intersensory redundancy that enables infants to bootstrap the information about the labeling (Bahrick & Lickliter, 2012; Gogate & Bahrick, 2001; Spelke, 1983). Such approaches appropriately point to the role of perception in various learning processes, but leave unanswered the question about infants' behaviors that cannot be explained by perceptual saliency alone, such as intermodal rule-learning (Kovács & Mehler, 2009a, 2009c) or learning words where the stimuli are matched for saliency (Smith & Yu, 2008). Similarly, these approaches do not explain in detail how infants recognize that events in different modalities are somehow connected. The inability to show that the intermodal integration is learned indicates that part of this knowledge has to be innate (Spelke, 1983).

The second direction of approaches to word learning concentrates on the role of domain-general learning processes. Associative learning – in which learning about a future event is based on the frequency or conditional probability of its occurrence in the past – is not limited to the domain of language (Fiser & Aslin, 2002; Gebhart, Newport, & Aslin, 2009), and has been shown to exist in several other species (Toro & Trobalón, 2005). The concern for domain-general learning mechanisms was re-created as a counterbalance to the view that proposed that language acquisition is governed by mechanisms – to a large extent innate – specific to language. It opened a new window for exploring in greater detail the role that these mechanisms have in language acquisition. Associative mechanisms were thus proposed to account for both word segmentation (Aslin et al., 1998; Saffran et al., 1996) and word learning (Smith & Yu, 2008). Some word segmentation studies show that non-statistical properties of language such as prosody can serve as a reliable cue for word segmentation and mapping (Endress & Hauser, 2010; Endress, Nespor, & Mehler, 2009; Millotte et al., 2010; Shukla et al., 2011). However, the majority of studies with preverbal infants assume that statistical computations form an important part of word segmentation: they either feed or underlie other cues for word segmentation (Thiessen, Hill, & Saffran, 2005; Thiessen & Saffran, 2003, 2007). Nonetheless, we miss data that would show how informative statistical segmentation really is in natural languages, and how infants use statistical information in statistically ambiguous situations.

The third line of researches emphasizes the role of innate language-specific knowledge in language acquisition. Whereas there is strong evidence suggesting that at least some parts of the acquisition of grammar are not conditioned by rich empirical experience (Chomsky, 1959; Lidz, Waxman, & Freedman, 2003; Tahakashi & Lidz, 2007; Wexler, 2013), it is less clear which innate universal properties characterize lexical acquisition. A large body of research was carried out on infants' speech perception, where infants were shown to react differently to linguistic and non-linguistic stimuli (Benavides-Varela et al., 2011; Nazzi, Bertoncini, & Mehler, 1998; Peña et al., 2003; Peña, Pittaluga, & Mehler, 2010). As for the more specific task of learning words in noisy natural environments, it was shown that syntax can guide the acquisition of novel words (Gleitman, 1990; Landau & Gleitman, 2009), and that the process of mapping words to correct referents may be guided by the language-specific principle of mutual exclusivity, according to which, by default, each referent has only one label (Halberda, 2003; Markman & Wachtel, 1988; Trueswell et al., 2013). Much of this work was carried out with toddlers, who already possess a substantial knowledge about their native language, so the question remains whether preverbal infants employ the same mechanisms or not.

The scope of this thesis is to question the role of associative mechanisms and perceptual saliency in word learning. We will assume that infants are able to reason about the events that surround them as soon as they possess some knowledge about the physical world, and that the nature of reasoning is not very different from that of adults (Renée Baillargeon, Li, Gertner, & Wu, 2011; Cesana-Arlotti, Téglás, & Bonatti, 2012; Téglás, Girotto, Gonzalez, & Bonatti, 2007). Perceptual saliency should thus not play an exclusive role in word learning. We will further assume that language processing is to some extent unconnected to other mental processes (Pinker, 1994). It is therefore unlikely that only general associative processes would account for language processing and learning.

In the first part of this thesis, we ask whether 4-month-old infants can represent objects and movements after a short exposure in such a way that they recognize either a repeated object or a repeated movement when they are presented simultaneously with a new object or a new movement. If they do, we ask whether the way they observe the visual input is modified when auditory input is presented. We investigate whether infants react to the familiarization labels and to novel labels in the same manner. If the labels as well as the referents are matched for saliency, any difference should be due to processes that are not limited to sensorial perception. We hypothesize that infants will, if they map words to the objects or movements, change their looking behavior whenever they hear a familiar label, a novel label, or no label at all.

In the second part of this thesis, we assess the problem of word learning from a different perspective. If infants reason about possible label-referent pairs and are able to make inferences about novel pairs, are the same processes involved in all intermodal learning? We compared the task of learning to associate auditory regularities to visual stimuli (reinforcers), and the word-learning task. We hypothesized that even if infants succeed in learning more than one label during one single event, learning the intermodal connection between auditory and visual regularities might present a more demanding task for them.

The third part of this thesis addresses the role of associative learning in word learning. In the last decades, it was repeatedly suggested that co-occurrence probabilities can play an important role in word segmentation. However, the vast majority of studies test infants with artificial streams that do not resemble a natural input: most studies use words of equal length and with unambiguous syllable sequences within word, where the only point of variability is at the word boundaries (Aslin et al., 1998; Saffran, Johnson, Aslin, &

Newport, 1999; Saffran et al., 1996; Thiessen et al., 2005; Thiessen & Saffran, 2003). Even if the input is modified to resemble the natural input more faithfully, the words with which infants are tested are always unambiguous – within words, each syllable predicts its adjacent syllable with the probability of 1.0 (Pelucchi, Hay, & Saffran, 2009; Thiessen et al., 2005). We therefore tested 6-month-old infants with such statistically ambiguous words. Before doing that, we also verified on a large sample of languages whether statistical information in the natural input, where the majority of the words are statistically ambiguous, is indeed useful for segmenting words. Our motivation was partly due to the fact that studies that modeled the segmentation process with a natural language input often yielded ambivalent results about the usefulness of such computation (Batchelder, 2002; Gambell & Yang, 2006; Swingley, 2005).

We conclude this introduction with a small remark about the term *word*. It will be used throughout this thesis without questioning its descriptive value: the common-sense meaning of the term *word* is unambiguous enough, since all people know what are we referring to when we say or think of the term *word*. However, the term *word* is not unambiguous at all (Di Sciullo & Williams, 1987). To mention only some of the classical examples: (1) Do *jump* and *jumped*, or *go* and *went*, count as one word or as two? This example might seem all too trivial, especially in languages with weak overt morphology as English, but in some languages, each basic form of the word has tens of inflected variables. (2) A similar question arises with all the words that are morphological derivations of other words, such as *evict* and *eviction*, *examine* and *reexamine*, *unhappy* and *happily*, and so on. (3) And finally, each language contains many phrases and idioms: Does *air conditioner* and *give up* count as one word, or two? Statistical word segmentation studies in general neglect the issue of the definition of words, assuming that phrases and idioms have strong internal statistics and will therefore be selected as one word (Cutler, 2012). But because compounds or phrases are usually composed of smaller meaningful chunks, it is unclear how would infants extracts these smaller units of speech if they were using predominantly statistical information. We will address the problem of over-segmentations shortly in the third part of the thesis.

2

Encoding and labeling discrete movements and objects at 4 months

2.1 Introduction

Current literature on early visual processing abilities indicates that infants can process physical reality by identifying individual objects by their outer shape (Spelke, Breinlinger, Jacobson, & Phillips, 1993; Van Giffen & Haith, 1984), by tracking their motion trajectories (Johnson, Amso, & Slemmer, 2003; Johnson, Bremner, et al., 2003), and by assigning them more basic or abstract properties such as continuity (Spelke, Katz, Purcell, Ehrlich, & Breinlinger, 1994), permanence (Baillargeon, 1987), invariance (Johnson, Amso, et al., 2003; Möhring & Frick, 2013; Moore & Johnson, 2011), color (Spelke et al., 1993), animacy, and agency (Luo, Kaufman, & Baillargeon, 2009). In general, these abilities emerge in the first moths after birth and are developed in detail during the first two years of life. Infants pay attention to different motions (Bahrick, Gogate, & Ruiz, 2002), track the object trajectory even if the object is occluded for a short amount of time (Gredebäck & von Hofsten, 2004, 2007; Elizabeth S Spelke & Kinzler, 2007), and remember them over time (Bahrick et al., 2002; Bahrick & Pickens, 1995). However, it has been generally assumed that the perception of a movement is closely related to the perception of the object that performs it, or even that movements constitute a part of information about the object (Leslie, Xu, Tremoulet, & Scholl, 1998; Xu & Carey, 1996).

The mature human brain processes visual information in two distinct neural pathways (Goodale & Milner, 1992). Whereas characterization of ventral stream as a “what” pathway, responsible for object recognition and representation, seems relatively uncontroversial, dorsal stream has been characterized either as a “where” or, more recently different “how” pathways, responsible both for spatial or action representation (Cloutman, Binney, Morris, Parker, & Lambon Ralph, 2013; Kravitz, Saleem, Baker, & Mishkin, 2011; Ungerleider & Haxby, 1994). Despite the more diverse functions of the dorsal stream and the possible unification of the two streams in higher cognitive representations, the two remain separate throughout the visual cortex (Cloutman et al., 2013). Experiments with congenitally blind participants suggest that representations in the two neural pathways do not require visual experience to mature but may be genetically endowed in the nervous system (Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009). And if this segregation is actually genetically endowed, there is an apparent gap between the current state of understanding of visual processing in adults (and non-human species) and in studies on infant visual abilities.

A similar gap can be found in studies of infants' labeling and reasoning abilities. Infants learn many different words till their second birthday. Word learning situations in everyday life are most commonly ambiguous, so infants often have to decide which word labels to assign to which referent – an object, an action or their parts. Given the complexity of the speech infants hear and of the visual scenes they see, the task should be almost impossible (Hochmann, 2013; Quine, 2013). But infants are surprisingly successful learners, and their referential errors are extremely rare (Bloom, 2001). To explain this paradox, there have been many principles and constraints proposed to work in favor of selecting the correct label for the right referent in ambiguous situations: besides various social cues that infants follow to disambiguate referents (Bloom, 2001; Csibra & Gergely, 2009), infants seem to have whole object bias when referencing (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Xu & Carey, 1996), they are prone to categorize based on labels they hear (Ferry et al., 2013), and are, importantly, able to pair novel labels with novel referents. To achieve the latter, there are at least three possible strategies they might use: cross-situational statistical learning (Smith & Yu, 2008; Yurovsky, Hidaka, & Wu, 2012), Novel-Name Nameless-Category Principle – the more perceptually based preference for novelty, which was suggested to explain why novel objects are labeled with novel labels (Golinkoff et al., 1994), or mutual exclusivity principle, according to which every referent has by default only one name and which might reflect a more general logical principle of disjunctive syllogism (Halberda, 2003, 2006; Markman & Wachtel, 1988).

Yet, all principles and strategies proposed to operate in early word learning were tested only with infants older than 12 months and only with labels for discrete objects, which would usually belong to the lexical category of nouns. Verbs and other categories are claimed to be learned later than nouns and with more difficulty (Gentner & Bowerman, 2009; Golinkoff & Hirsh-Pasek, 2008; Maguire et al., 2010; Maguire, Hirsh-Pasek, & Golinkoff Michnick, 2006). And as for word learning early in life, it is usually claimed that infants in their first year of life lack the reasoning and social abilities necessary to learn words in an adult-like manner (Bloom, 2001). Some studies have shown that infants as young as 6 months are able to extract a label from speech and map it onto a referent only in an unambiguous situation where the label, the object, and the connection between them are all salient (Bahrick & Lickliter, 2012; Gogate & Bahrick, 1998, 2001; Shukla et al., 2011), but adult-like learning in ambiguous naming situations has not been shown in a laboratory environment in infants prior to 12 months.

However, it has recently been shown that infants as young as 6 months come to the laboratory already knowing some of the most common words (Bergelson & Swingley, 2012). They have most probably learned them in a noisy environment where they had to choose the most probable referent, just like their older fellows. It is therefore most plausible that infants possess all the necessary reasoning abilities a few months after birth.

In light of the contradictions described above, the aim of this chapter is to show that 4-month old infants already posses the abilities to represent objects and movements as two separate entities, that they can label both objects and movements when the labeling situation is salient, and that these abilities resemble those of older infants and adults in many important aspects. The first part of Chapter 2 will describe experiments in which the ability to rapidly encode and represent both objects and movements is assessed in 4-month-old infants. The second part describes experiments in which we examined whether infants change the way they observe the visual scene when the labels for objects and movements are present. In both parts, the "rapid visual recognition" testing procedure (RVR) was used. RVR is a novel procedure built on the assumption that we should be evolutionary endowed to rapidly observe our environment and that normally developing infants should therefore need a short time to scan visual scenes and to individuate or recognize salient parts of the scene (Öhman, 1997). In many experiments about perception and/or recognition (and not about learning), long exposures are therefore not needed and might even hinder results due to drops of attention. In RVR, each trial has a very short familiarization phase with one item in the center of the screen, followed by a short test phase with two simultaneous items, each at one side of the screen. One item is a (partial) repetition of the familiarization and the other is novel.

2.2

Representation of movements and objects at 4 months

The first two experiments address the question of stable or independent representation of both objects and movements in 4-month-old infants. In the first experiment, we examined whether infants recognized a repeated object, moving in a different direction, and a movement, performed by a different object, after a short exposure to a moving object. In the second experiment, we tested how stable the representations of objects and movements after a short exposure is by rotating both movements and objects.

2.2.1 Experiment 1

Recognition of movements and objects

2.2.1.1 Methods

2.2.1.1.1 Subjects

In Experiment 1 we tested 34 infants. The final analysis contains the looking-behavior of 16 infants who finished the experiment (9 boys, 7 girls, mean age 137 days, range 106 to 157). 18 infants were rejected from the analysis because of fuzziness or unsuccessful calibration, or because less than 50% of the total possible looking-time samples could be collected with the eye-tracker during the experiment. The majority of rejected infants were younger than 105 days (3.5 months). At that age keeping an upright posture and controlling head position is still underdeveloped in some infants, and the pupil and cornea contrast may not be sufficient. Given the fact that infants are tested sitting in the lap of one of the parents (see Procedure section below), the final group of accepted babies constitutes the youngest age at which we in our lab were able to test infants with the eye-tracker. All infants were born without complications and none of the parents reported their infant having visual or auditory problems. The APGAR values for all infants were 8 and above. Informed consent was obtained from the parents of all infant participants before the beginning of the experiments. The infant experiments were approved by the Bioethics Committee of SISSA – International School for Advanced Studies (date of the approval 25/03/2010). Parents received a small monetary compensation for travelling expenses.

2.2.1.1.2 Materials

The stimuli of Experiment 1 consisted of 8 abstract two-dimensional shapes controlled for color, texture, size and other low-level visual cues (e.g. curvature vs. linearity). The objects could move in 8 different directions separated by 45° angles (always from the center outward, see Figure 1), with different movements being controlled for distance from the center of the frame and the speed with which the shape moved. This resulted in 64 different videos, each video consisting of 3 repetitions of the same moving object on a black background (total length 2400 ms; frame size 600x600 px; frame rate 60 fps). The videos were used both in the familiarization and the test phases of each trial. In total, each of the objects and each of the movements were presented 6 times.

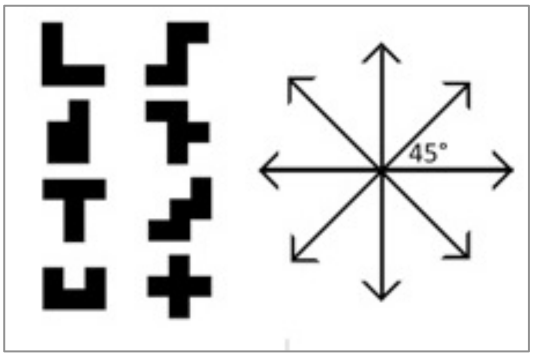

Figure 1 Materials in Experiment 1.

2.2.1.1.3 Procedure

Infants were presented with Object recognition (n=8) and Movement recognition (n=8) trials in random order. Each trial contained a familiarization video followed by two simultaneous test videos, one on each side of the screen – a video with either familiarization object or movement repeated and a novel combination of a movement and an object. Presentation of novel and familiar objects was side-balanced across trials. Familiarization and test sequences were separated by 1200ms-long visual silences because previous studies have reported that shorter intervals could induce apparent movement effects (Gredebäck & von Hofsten, 2007), or infants could think the objects as continuous (S. P. Johnson, Amso, et al., 2003). Total trial duration was thus 6000ms. The trials were separated by central fixations and the experimenter initiated each trial when the infants' gaze was directed to the center of the screen (Figure 2).

Infants' gaze was recorded with a TOBII T60 eye-tracker (http://www.tobii.com/) at a rate of 60 Hz. The eye-tracker was integrated into a 17-inch TFT screen. Before the start of the experiment, participants' eye-movements were calibrated with Tobii Studio software. We used a five-point calibration procedure in which they followed an attention-grabbing blue looming ball that moved to the four corners of the screen and then to the center of the screen. To attract additional attention, if necessary, we used a yellow duckling that appeared randomly on the screen. If fewer than 4 points were calibrated on any of the two eyes, the procedure was repeated. The participants typically required one or two calibrations. The experiment immediately followed the calibration procedure. The stimuli of the experiment were presented via PsyScope X software (http://psy.cns.sissa.it/). Gaze recordings were segmented and grouped using Perl program; statistical analysis was run with Matlab 7.9.0 software.

Infants were seated on their parent's lap (in order to avoid any discomfort due to a new environment without a presence of a familiar person) at about 60cm distance from the monitor. Parents wore blocked glasses to avoid the eye-tracker collecting their gaze and to ensure that infants' looking-behavior was not affected by parental influence. To determine whether infants' looking behavior to the two test videos in each trial differed, we determined a region of interest that matched the size and the location of the videos on the screen (600x600 pixels). Only the looks that fell into these ROIs were counted in the measures of looking times for each of the two test videos. We measured cumulative looking time to each of the ROIs (the sum of all the looks during the test phase), the longest uninterrupted look to each of the ROIs, and the first fixations during the test phase (eg. at least 100ms of uninterrupted time spent in a ROI).

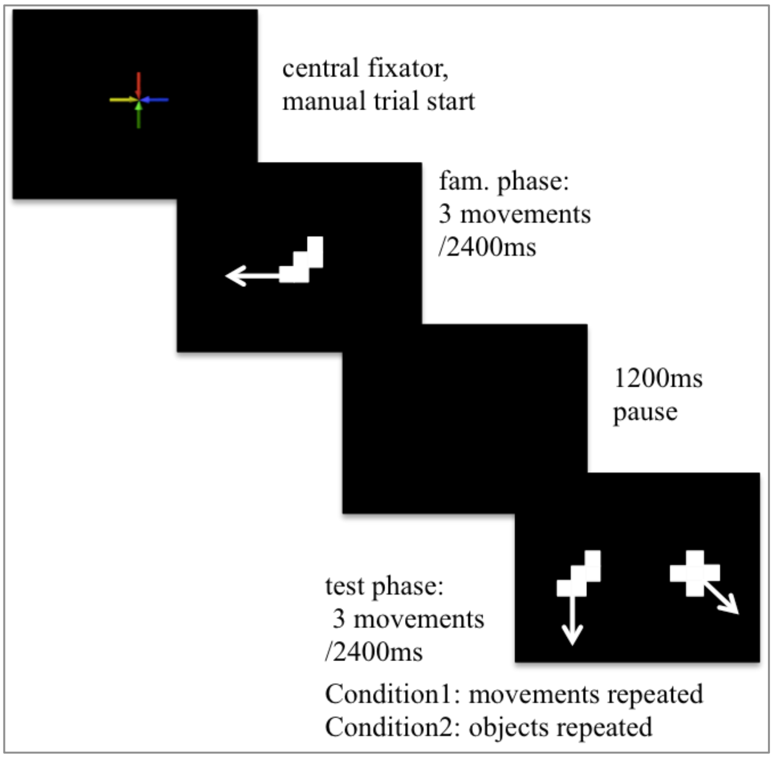

Figure 2 Trial structure in Experiment 1. The trial presented in the Figure represents an Object recognition trial (Condition 2). Movement recognition trials (Condition 1) were equal, except that the Test phase contained two new objects, one moving in the same direction as the object in the Familiarization phase. White arrows represent the direction of movement; in all the movements, the objects moved from the center to the extremities of the frame and then reappeared in the center.

2.2.1.2 Results

The results of Experiment 1 are shown in Figure 3. For each infant we computed the difference score (d) of the difference between total amount of looking to the novel (n) and total amount of looking to the familiar (f) videos (d=(n-f)/(n+f)) across the trials in each condition (Edwards, 2001). The span of the difference score was therefore from 1.00 (novelty preference) to -1.00 (familiarity preference), where 0 represents no preference. All the reported effect sizes are computed with Cohen's *d* (the difference between the means/pooled standard deviation). Infants spent on average significantly more time looking at novel events than at events that contained either the object or the movement seen during familiarization (ME(d)=0.253; SD=0.303; paired 2-tailed t-test against 0: t(15)=3.341, p=0.004; Cohen's *d*=1.18). Significant novelty preferences in cumulative looking-times prevailed in both Object (ME(d)=0.286, SD=0.465; t(15)=2.456, p=0.027;

Cohen's *d*=0.868) as well as Movement (ME(d)=0.247, SD=0.385; t(15)= 2.564, p=0.022; Cohen's *d*=0.906) recognition trials. Infants were also likely to direct their longest uninterrupted looks at events containing movement-object combinations they had not seen before (ME(d)=0.223, SD=0.237; paired 2-tailed t-test against chance level: t(15)= 3.418, p=0.0038; Cohen's *d*=1.208). For the longest fixations, too, significant novelty preferences prevailed in both Object (ME(d)=0.265, SD=0.421; t(15)=2.516, p=0.024; Cohen's *d*=0.889) as well as Movement (ME(d)=0.180, SD=0.274; t(15)=2.622, p=0.019; Cohen's *d*=0.927) recognition trials. In Object recognition trials, infants directed their first fixation to the novel object marginally significantly (ME(d)=0.273, SD=0.555; t(15)=1.974, p=0.067; Cohen's *d*=0.698); when only the movement was repeated in Movement recognition trials, they showed no such preference (ME(d)=-0.009, SD=0.365; t(15)=0.103, p=0.919; Cohen's *d*=0.036). Across participants, there was no significant difference between looking behavior in Movement and Object repetition trials (paired two-tailed t-test between the difference scores of cumulative looking times in both conditions: t(30)=0.256, p=0.799; Cohen's *d*=0.09).

Because of the possibility that the results were partially driven by the preference for a particular object or a particular movement, we performed the analyses of variances between the average looking times to each of the objects and to each of the movements. During the test phases, infants saw each object and each movement 4 times, for 9600ms in total. The average looking time to each object and movement was 2.755ms (SD=450ms). The one-way analysis of variance between average looking time to each of the 8 objects showed no significant preferences for individual objects (F(7,120)=1.606, p=0.140). Similarly, no differences were found between looking time spent on each of the 8 movements (one-way ANOVA: F(7,120)=0.752, p=0.628). These consistent novelty preferences could thus only emerge if infants recognized the repeated objects/movements in one of the test events.

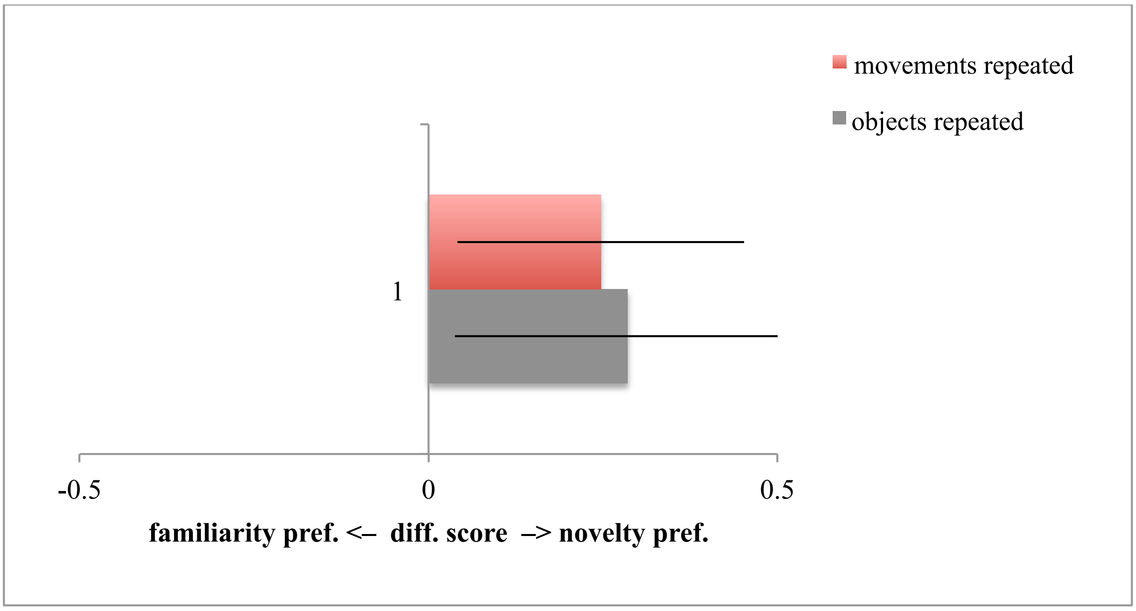

Figure 3 Results of Experiment 1. The bars represent difference score (d=(n-f)/(n+f)) of total looking time to either of the sides of the screen in each of the two conditions. The error bars represent 95% confidence intervals.

2.2.1.3 Discussion

The results of the Experiment 1 confirmed our hypothesis that infants can rapidly encode and represent objects and thus recognize the repetition of an object after a short pause, focusing more to the object that they have never seen before. When the test phase contained no repeated objects but only one repeated direction of movement, they recognized this abstract repetition and again focused on the novel combination of object and movement, indicating that representations of movements are separate from representations of objects. The fact that there are no quantitative differences in looking time across both conditions offers a further confirmation to the hypothesis that they are equally fast at recognizing objects independently of their movements, and movements independently of the object performing it.

These results could however represent a more elementary ability to recognize repetitions of any kind, attested in very young infants both in visual and auditory stimuli (Endress, Scholl, & Mehler, 2005; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; Matuz et al., 2012). We therefore conducted another experiment in which we asked whether the representation of objects is stable enough so that they recognize an object after a short exposure, even if it is rotated. In second half of the trials we also rotated movements according to the rotation of the object, to address the question whether infants perceive movements as a feature of the object, or as independent representations.

2.2.2 Experiment 2

Recognition of rotated movements and objects

In natural environment, objects are recognized when they change location, and even when they change the angle of view, when they are rotated or turned upside down. There are various hypotheses about how are these different projections recognized as one single object: objects may be either represented in memory as structural descriptions in object-centered coordinate systems, so that the representation is identical regardless of its orientation (Marr & Nishihara, 1978), or they could be represented in memory in a single canonical orientation, while other representations are recognized via the process of rotation (Tarr & Pinker, 1989). It has been shown that infants can recognize rotated objects (Moore & Johnson, 2008). The question in our experiment is whether infants represent movements as a feature of the object, and whether the representation of the movement can be rotated together with the object.

2.2.2.1 Methods

2.2.2.1.1 Subjects

In Experiment 2 we tested 42 infants. The final analysis contains the data of 24 infants who finished the experiment (8 girls, 8 boys, mean age 120 days, range 108 to 131). 18 infants were rejected from the analysis because less than 50% of the total possible looking-time samples could be collected with the eye-tracker during the experiment (because they didn’t finish the experiment or because of the eye-tracker failure). Other characteristics of the subject sample are the same as in the Experiment 1.

2.2.2.1.2 Materials

The stimuli of Experiment 2 consisted of same abstract two-dimensional white shapes used in Experiment 1, 4 of them used predominantly and 4 of them used only in the Object rotation trials. These objects could move in 4-different directions separated by 90° angles from the center outwards. The videos of these moving objects were then rotated to 90°, 180° and 270°. The videos’ length and size was the same as in the Experiment 1. In total, infants saw each movement 8 times and each of the predominantly used objects 10 times (and the remaining ones 2 times each, in Object rotation trials). The predominantly used objects and the movements are depicted in Figure 4.

2.2.2.1.3 Procedure

Infants were presented with Object rotation (n=8) and Movement rotation (n=8) trials in random order. Each trial contained a familiarization video followed by two simultaneous test videos. In the test videos of Object rotation condition, infants saw a new object and the previously seen object rotated. In the Movement rotation condition, infants saw two equal objects on the screen, identical to those in the familiarization phase but rotated, one moving in a new direction, the other moving in the direction rotated according to the rotation of the object (so that the whole event was rotated, not only the object; see the object depicted on the right side in the test phase of Figure 4). The trial structure, duration of the experiment, presentation, calibration, recording, coding of the gazes, and statistical analysis were the same as in Experiment 1.

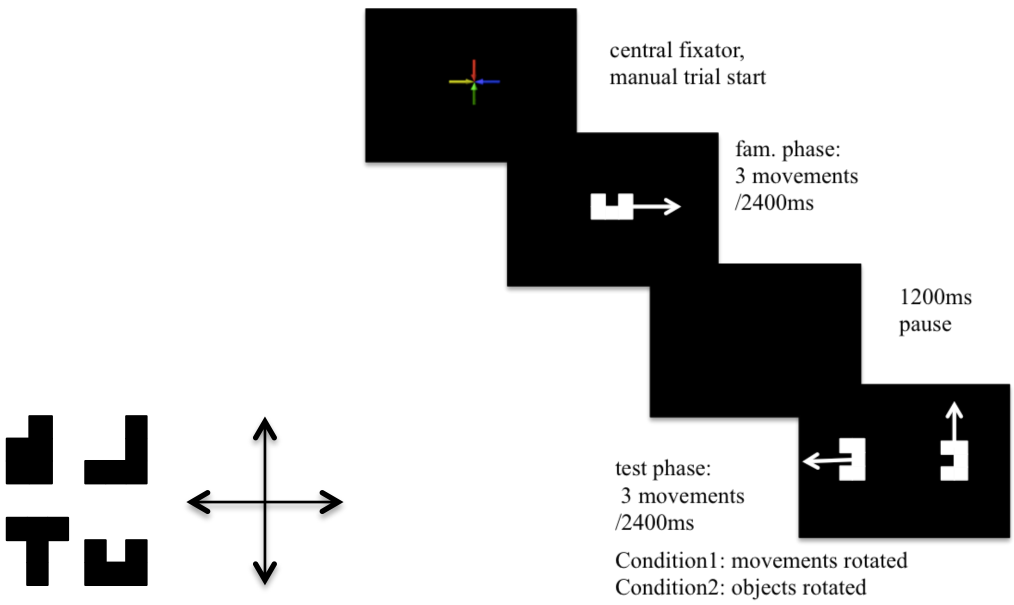

Figure 4 Materials and trial structure in Experiment 2. Trial presented at the Figure represents a Movement recognition trial (Condition 1). In Object recognition trials (Condition 2) they saw a repeated object rotated and a novel object, both moving in new directions. White arrows represent the direction of movement.

2.2.2.2 Results

The results of Experiment 2 are shown in Figure 5. As in Experiment 1, normalized difference scores were computed to assess novelty preference during the test phase. During the Object rotation trials, infants on average spent significantly longer time looking to the novel object although the repeated object was rotated and moved in a new direction (ME(d)=0.237, SD=0.313; paired two-tailed t-test against 0: t(23)=3.033, p=0.008; Cohen's *d*=1.072). Angle-based analysis revealed differences among the trials according to the angle of rotation. Infants recognized objects and looked more to the novel ones when rotation angles were 90º (ME(d)=0.488, SD=0.331; paired two-tailed t-test against 0: t(23)=5.897, p=0.00003) and 270º (ME(d)=0.332, SD=0.526; paired two-tailed t-test against 0: t(23)=2.521, p=0.023), but not when the angle was 180º (ME(d)=-0.115, SD=0.720; paired two-tailed t-test against 0: t(23)=-0.642, p=0.531). In fact, infants' gaze behavior differed significantly according to the angle of rotation, as indicated by the analysis of variance between the average looking times in each of the three rotation angles (one-way ANOVA with the factor Angle: F(2,45)=5.202, p=0.009). Post-hoc multiple comparison in Matlab revealed that there was no significant difference in looking behavior when the objects were rotated to 90º and 270º, but that the behavior was significantly different when the angle of rotation was 180º (Figure 6).

During Movement rotation trials, infants saw two equal objects on the screen, one of them moving in a novel direction, while the other was moving in the direction rotated for the same angle as the object (i.e. the whole scene was rotated). Infants did not fixate any longer the movements and showed no overall preference for any of the two events on the screen (average difference score: ME(d)=0.05, SD=0.309; paired two-tailed t-test against 0: t(15)=-0.751, p=0.464). Angle-based analysis of the differences between the rotation angles of the movements revealed no significant differences among the looking times to different axes of the movement (one-way ANOVA with the factor Angle: F(2,45)=1.249, p=0.296).

Based on the suggestions that gender difference in spatial perception is present in humans from early infancy – in a series of recent studies, boys recognized rotated objects whereas girls did not (Moore & Johnson, 2008, 2011) – we measured gender differences among the looking times to the novel object also in our experiment. However, we found no significant differences among boys and girls in total amount of time they spend looking to the novel object (two-way ANOVA with factors Angle and Gender revealed the expected significant effect of Angle ($F(2,45)=5.098$, $p=0.010$) and no effect of Gender ($F(1,45)=2.064$, $p=0.158$). There was no interaction between the two factors ($F(2,45)=0.016$, $p=0.984$).

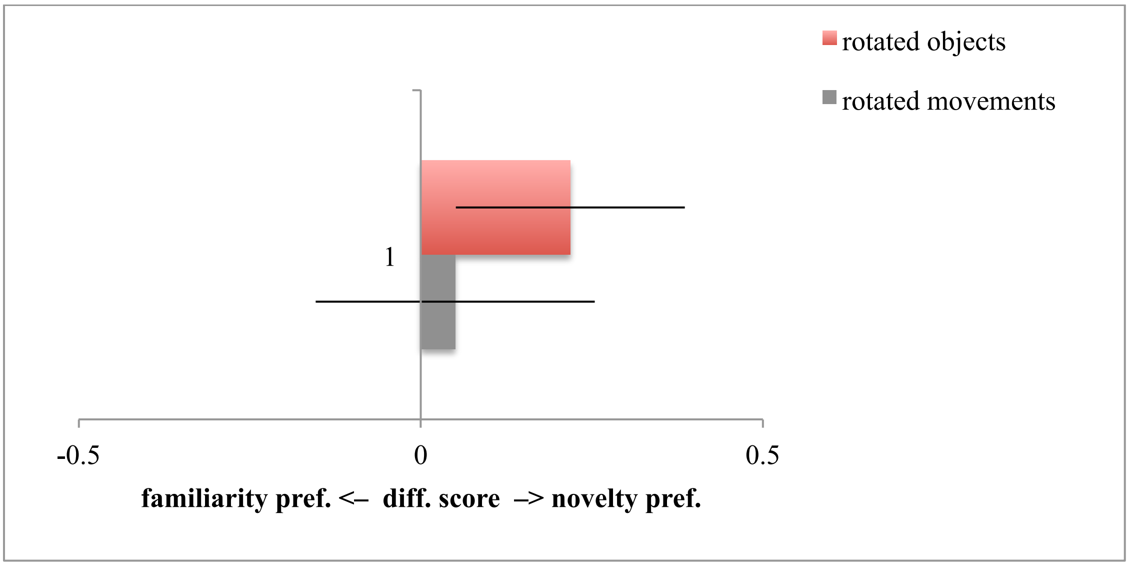

Figure 5 Results of Experiment 2. The bars represent difference score of total looking time to either of the sides of the screen in each of the two conditions. The error bars represent 95% confidence intervals.

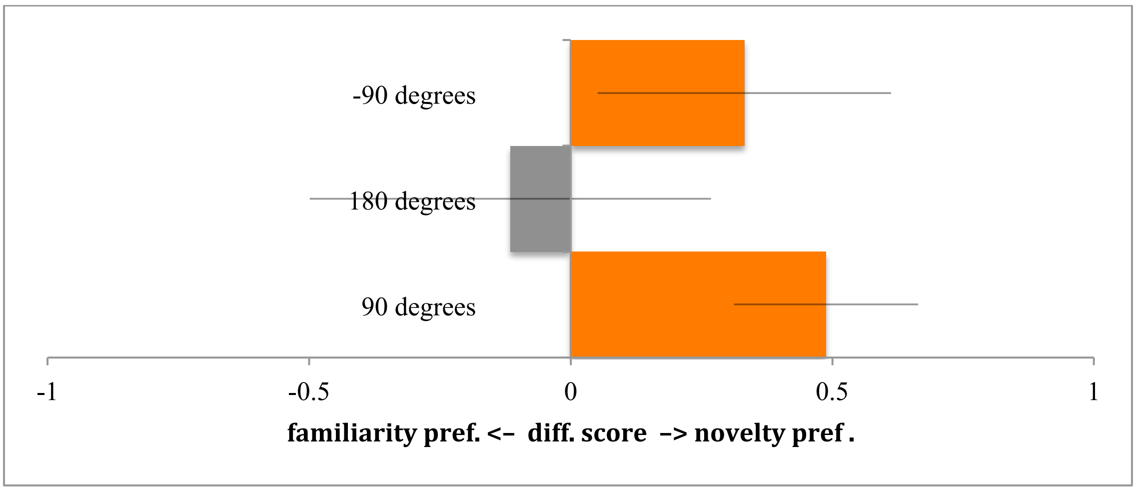

Figure 6 Angle-based analysis of the responses during Object rotation trials.

2.2.2.3 Discussion

In Experiment 2 infants recognized repetitions of objects, as in Experiment 1, but did not recognize the movements rotated according to the rotation of the object in the Movement rotation trials. More specifically, infants recognized repeated objects if they were rotated in different angles. The results confirm our hypothesis that infants rapidly represent objects they are exposed to and recognize them after a short pause, even if the viewing angle is different. These results further support results from other studies in that objects are easier to recognize if they are rotated only slightly (in our experiment 90º), and more difficult if they are rotated half way through the full rotation circle, 180º (Tarr & Pinker, 1989). Thirdly, these results seem not to replicate previously attested gender differences in spatial perception (Moore & Johnson, 2008, 2011); infants in our study show no gender difference in their ability to recognize rotated objects.

Moreover, the failure to recognize rotated movements in the Movement rotation trials could be predicted if infants really represent movements as independent from the objects that perform the movements: if the direction of the movement does not constitute an integrated part of the information about the object, then the rotation of the object should not in any way predict the direction of the movement. The results during the Movement rotation trials confirm this hypothesis. There is however an unrelated issue that can be raised regarding the perception of movements in this experiment. Infants are sensitive to the direction in which visual stimuli are presented (Bomba & Siqueland, 1983). It is thus possible that infants in our experiment recognize some movements as more familiar than others because they group the directions of the movements into general paths, such as vertical vs. horizontal, and that such a grouping plays a role in the processing of movements. Due to the unbalanced number of trials in which the vertical and horizontal paths are confronted in this experiment, a control study would be needed to answer this question.

The results of the first two experiments indicate that infants indeed process their visual input more rapidly that previously thought and that the visual processing of movements and objects is categorically different in infants. In order to address the question of whether this categorical distinction can be extended to labeling situations, we created two experiments in which infants saw the same visual material, but the labels for each object and each movement were added. We asked whether infants map new labels to new objects and what are the possible mechanisms that guide the mapping.

2.3

Labeling movements and objects at 4 months

Experiment 3 addresses the question of rapid labeling for objects, and Experiment 4 labeling of movements. We used a very similar procedure as in Experiment 1, and the same set of visual stimuli, which infants show they are able to discriminate in a short period of time. Adding the labels to the same set of stimuli enabled us to observe whether the perception of visual stimuli changes when the labels are added. Additionally, it enabled us to observe how do infants react to labeling without a long familiarization procedure and without the need to remember the labels for more than a few seconds.

2.3.1 Experiment 3

Labeling objects

2.3.1.1 Methods

2.3.1.1.1 Subjects

In Experiment 3 we tested 36 infants. The final analysis contains the data of the 24 infants who finished the experiment (12 boys, 12 girls, mean age 132 days, range 112 to 150). 12 infants were rejected from the analysis because they didn't complete the task (n=9), had a strong bias to look to one side of the screen (n=2), or because of the eye-tracker failure (n=1). The criterion for the exclusion based on the side-bias was if the participant spent more than 75% of the time looking on one side of the screen. Other characteristics of subject sample are the same as in Experiment 1.

2.3.1.1.2 Materials

The visual material was the same as in Experiment 1, except that the duration of exposure differed. In the familiarization phase, infants only saw one repetition of the movement

(900ms), and during the test phase they saw two simultaneous objects; each moved 4 times (3200ms of exposure). Auditory material consisted of nonce words recorded by a female native speaker of Italian and normalized for duration and intensity (final word length was 800ms and intensity 70dB). All the words were bi-syllabic and stressed on the first syllable; all the syllables had the consonant–vowel (CV) structure. In order to construct contrastive stimuli, we balanced the number of different consonants and vowels used to construct the words. To control for the possible familiarity with any of the nonce words, we used two separate lists of 8 words (List A and B), randomly assigned to the participants (Table 1). Labels from each language were randomly assigned to the 8 objects. The relationship between an object and a label was systematically maintained throughout the experiment.

LIST A	LIST B
'BADE	'DAFE
'GHIFO	'BIGO
'DAKU	'ZENO
'LERO	'LIME
'NUPI	'PUNA
'KUMA	'RASI
'SATE	'TEVO
'VUDA	'GUPA

Table 1 Words that were used during Experiment 3. Participants were randomly assigned to listen to one of the two lists. The apostrophes mark the stressed syllables.

2.3.1.1.3 Procedure

Infants were presented with No label (n=8), Familiar label (n=8), and Novel label (n=8) trials in random order. The trial structure is presented in Figure 7. In each trial infants saw a familiarization video and simultaneously heard the label for the object presented, followed by two simultaneous test videos, one with the familiar and one with a novel object. Each of the three objects moved in different directions. In the No label condition, there was silence during the test phase; in the Familiar label condition they heard the same label as in the familiarization phase; and in the Novel label condition they heard a novel label. During the familiarization phase, each label-object combination was introduced only once in each condition, and in the test phase, each of the objects was paired with the same new object in all conditions, creating a conditional probability of 0.5 for each object co-occurrence. Each of the objects and each of the movements were thus presented 9 times during the experiments. The presentation of novel and familiar objects was side-balanced across trials. Familiarization and test sequences were separated by 1200ms-long visual silences. Trials were separated by central fixations and the experimenter initiated each trial when the infant's gaze was directed to the center of the screen. Infant's gaze was recorded with a TOBII T60 eye-tracker (http://www.tobii.com/) at a rate of 60 Hz. The eye-tracker was integrated into a 17-inch TFT screen. Stimuli were presented via EventIDE software (http://www.okazolab.com). The calibration procedure, handling infants, coding gazes, and statistical analysis was the same as in Experiments 1 and 2.

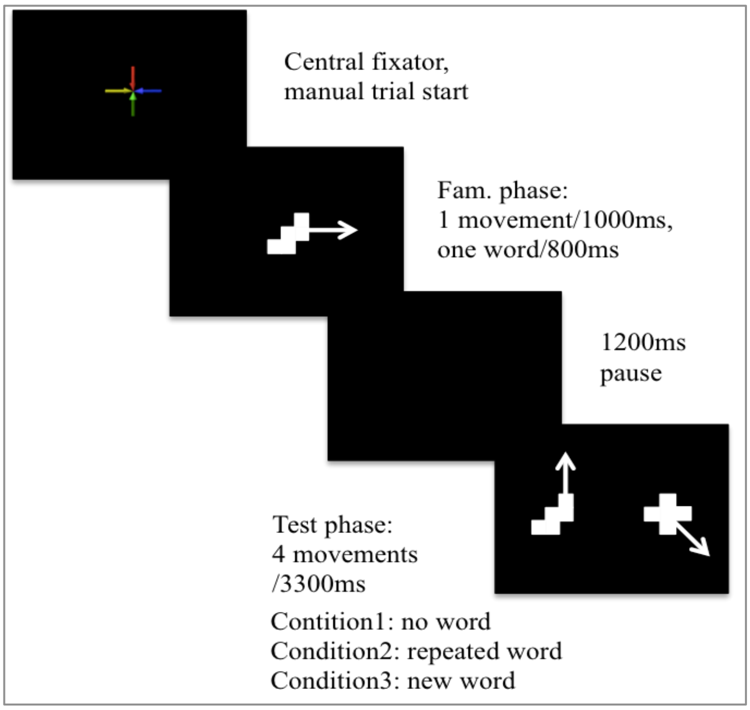

Figure 7 Trial structure in Experiment 3. The bars represent the direction of movement.

2.3.1.2 Results

In the test phase, infants on average started 63.3% (SD=16.9) of the trials by looking either to the center of the screen or away from the screen, and not to one of the sides of the screen where the test videos appeared (paired two-tailed t-test against chance level (50%): t(23)=3.877, p=0.0007). The first fixations that followed were not significantly directed to any of the two sides of the screen where the videos were presented (ME(d)=-0.071, SD=0.242; paired two-tailed t-test against 0: t(23)=-1.447, p=0.161, Cohen's *d*=0.410). After hearing the second label, infants however adapted their behavior according to the label that they heard.

The results of the total looking time during the test phase in each of the 3 conditions are presented in Figure 8. To assess the overall novelty preference, we computed the difference scores. The computation was the same as in Experiment 1. During the test phase of Condition 1, when there was silence, infants showed no overall preference for either familiar or novel objects (ME(d)=-0.012, SD=0.313; paired two-tailed t-test against 0: t(23)=0.183, p=0.856; Cohen' *d*=0.063). In Condition 2, after they heard the repeated word, infants on average spent more time looking at the familiar object (ME(d)=-0.164, SD=0.339; paired two-tailed t-test against 0: t(23)=-2.366, p=0.027 Cohen' *d*=0.683). And after hearing the novel word in Condition 3, they looked significantly more at the novel object (ME(d)=0.176, SD=0.351; paired two-tailed t-test against 0: t(23)=2.459, p=0.021; Cohen' *d*=0.710). The analysis of variance between the three conditions revealed that infants behaved significantly different in the three conditions (one-way ANOVA with the factor Condition: F(2,69)=5.44, p=0.006). Post-hoc multiple comparison in Matlab revealed significant differences in looking behavior between Conditions 2 and 3 (familiar vs. novel word), but no significant differences between these two groups and Condition 1 (no word).

Infants heard two different sets of words (12 infants heard the List A and 12 the List B), but there is no significant difference between the two subgroups of subjects (two way ANOVA with factors Condition and List showed no effect of List (F(1,66)=2.301, p=0.134), the expected significant effect of Condition (F(2,66)=5.411, p=0.006), and no interaction between the two (F(2,66)=0.171, p=0.843).

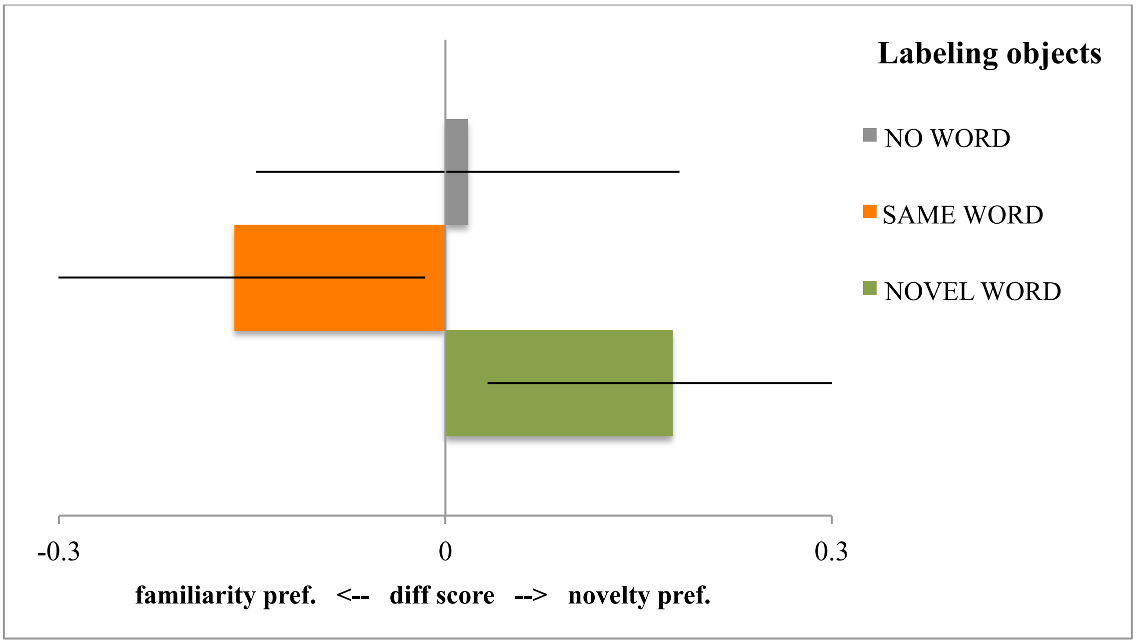

Figure 8 The results total looking time after hearing the second word till the end of the test phase of Conditions 2 and 3 (and the matching time-window in Condition 1) in Experiment 3. The error bars represent 95% confidence intervals.

During the course of the test phase, infants on average switched their gazes between the two simultaneous videos 1.928-times (SD=0.431). The number of the switches between the two sides of the screen did not differ across conditions (the analysis of variance with the factor Condition: F(2,69)=0.062, p=0.940). However, the majority of the infants switched their gazes to the familiar video if their first fixation happened to be on the novel video in Condition 2, when they heard the familiar label (ME=64.44%, SD=31.69, paired two-tailed t-test against 50% chance level: t(23)=2.233, p=0.036); similarly, infants switched their gazes to the novel video when hearing the novel word in Condition 3 if their first fixation was on the familiar video (ME=72.36%, SD=28.0, paired two-tailed t-test again 50% chance level: t(23)=3.913, p=0.0006).

To obtain more detailed information about the looking behavior of infants, we conducted a sample-based analysis of looks during test phase. For each recording sample in each condition, we averaged looking behavior across participants. The lines in Figure 9 represent looking behavior in all three conditions. Infants' looking behavior started to diverge immediately after hearing the first syllable of the word: when hearing the repeated word, they on average turned their gaze more to the familiar object (Condition

2), and when hearing the novel word, they switched their gazes to the novel object (Condition 3). In Condition 3, infants looked significantly more towards the new object immediately after hearing the word. Then the novelty preference decreased around 1000ms after the offset of the word, after which the gazes are again more significantly directed towards the novel object. In order to identify the points in time during the test phase when infants' average behavior differed significantly across different conditions, we performed the permutation test between the looking behavior in each data sample in Conditions 2 and 3 (we excluded Condition 1 because the overall behavior in Condition 1 did not significantly differ from 0 or from Conditions 2 and 3). We used a standard permutation-based t-test with 10,000 permutations to calculate the mean and standard deviation of the overlap. We assumed a normal distribution, with the p-values lower than 0.05 called significant. The results of the permutation test are represented in the background picture of Figure 9, which shows the data-points in which the results from Conditions 2 and 3 differ significantly. Grey shades represent the significance levels of the differences between the two conditions. The results show that the looking behavior started to differ significantly around the offset of the word (800ms) and remained significantly different during most of the test phase.

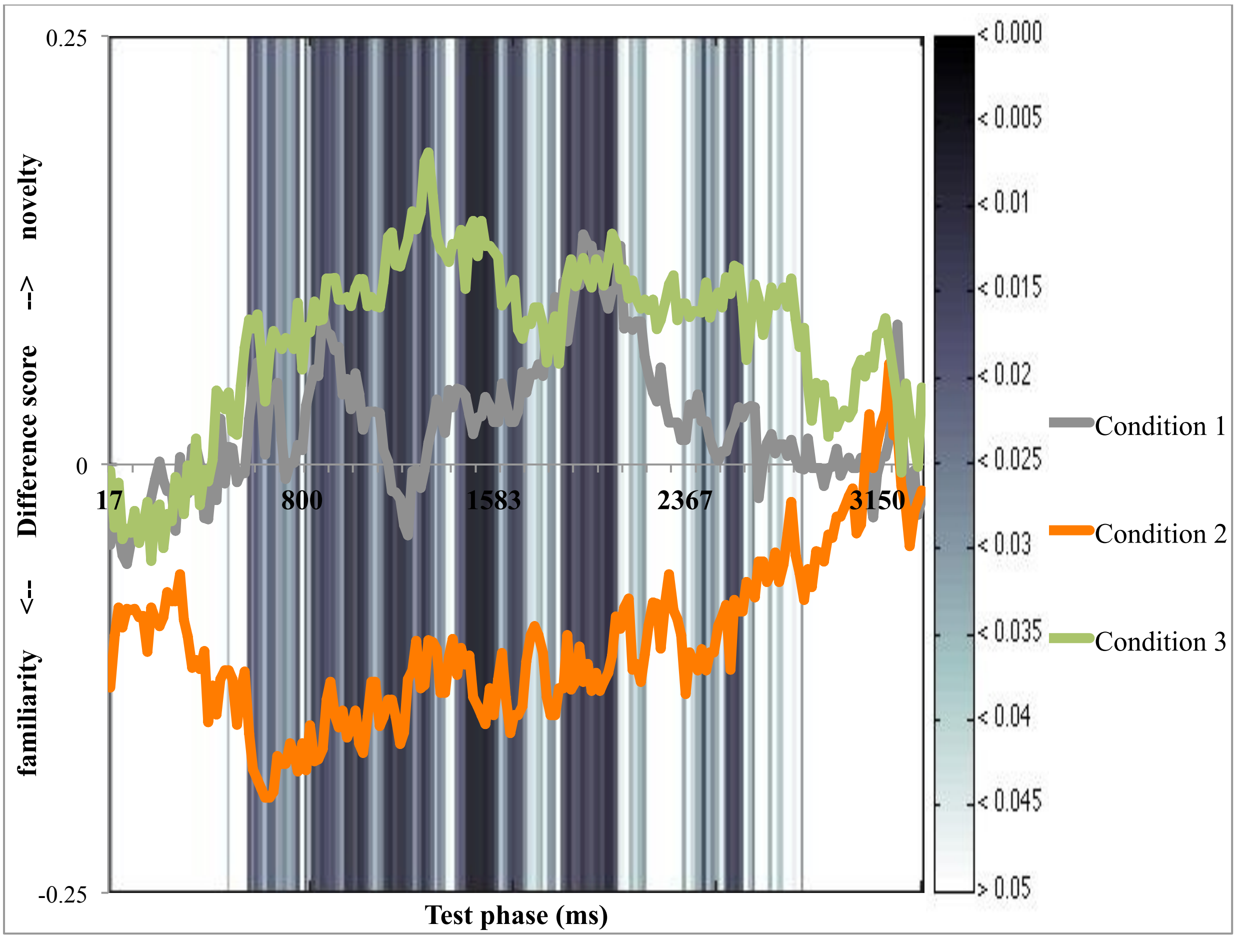

Figure 9 Sample based analysis of looking behavior in each of the three conditions during the test phase in Experiment 3. Positive scores signify looking at the novel object, and negative scores looking at the repeated object. Background figure represents the results of the permutation test between looking patterns in Condition 2 and 3. Grey shades represent the significance levels of the differences between the two conditions (as scaled in the color bar). The offset of the word in the Conditions 2 and 3 happened at 800ms.

2.3.1.3 Discussion

In this study, infants do not show any specific preference for either repeated or novel item when they hear no label in the test phase (Condition1). In Experiment 1 and 2 of this chapter, in which they received only visual stimuli, infants looked longer to novel items, both when the objects and when the movements were repeated (Langus, Saksida, & Nespor, under review). What could cause the change in the looking behavior in infants when they, on top of seeing a moving object, hear a label in the familiarization phase? It has been recently shown that infants as young as 10 months create referential expectations when hearing words (Csibra, 2010; Gliga & Csibra, 2009). We suggest that something similar happens in infants as young as 4 months: labeling in the familiarization phase in our experiment creates a referential expectation that overrides the novelty preference shown in the Experiments 1 and 2.

During trials with label repetition (Condition 2), infants look longer to the repeated item when they hear a repeated label. Because the only visual item repeated in the test is the object, they seem to infer that a mapping should exist between the repeated label and the repeated object, even after a single very short exposure (Spiegel & Halberda, 2011). Moreover, infants look longer to the novel item when they hear a new label, different from that of the familiarization phase (Condition 3). Whereas their first gazes were not significantly directed to any of the sides at the beginning of the test phase, they have alternated the behavior after hearing the labels, indicating that it was the labels that created this difference and not their processing of the visual scene.

Quite surprisingly, infants' recognition of the objects and labeling were successful even if the familiarization in this experiment was shortened to one single movement and the total exposure of 900ms. Adults can recognize an object if they see it for more than 20ms (Thorpe, Fize, & Marlot, 1996). Very few studies have been done with infants on the speed of visual processing, and there are indications that infants can recognize the objects after the exposure of more than 100ms (Kouider et al., 2013). While adults can consciously access the presented objects after ca. 300ms, in 5-month-old infants the process seems to take longer, up to 900ms. However, vast majority of studies on object recognition and on word learning uses much longer exposure times. This experiment thus confirms that both building object representations and object labeling are very rapid, probably to some extent automatic processes.

It is much less clear which mechanisms underlie such behavior. The experiment was designed to resemble the situations in which language learners have to make inferences about which word refers to which object. A principle that is commonly understood to help language learners in this task is called Mutual Exclusivity and refers to the fact that in everyday situations words are usually used unambiguously: each word has a different referent. If a novel word is encountered together with a novel referent, language learners tend to map the two (Halberda, 2003; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003; Spiegel & Halberda, 2011). It is predominantly used to explain the process of word learning, indicating that it might be a domain-specific principle used by human language learners only. But there are some recent studies that suggest that principle may not be species-specific, as indicated by the experiments with dogs (Bloom, 2004; Kaminski, Call, & Fischer, 2004; Markman & Abelev, 2004). It might not even be language-specific since it can apply to non-linguistic sound-object pairs such as the association of a voice to a specific face. Therefore, there have been some attempts to explain the principle of mutual exclusivity in terms of a domain-general logical principle of disjunctive syllogism (Halberda, 2003, 2006). Another very frequent explanation for the behavior in which language learners map novel labels to novel objects is that it is a consequence of a more perceptually based preference for novelty. Such explanations will assume something like Novel-Name-Nameless-Category Principle (Golinkoff et al., 1994; Mather & Plunkett, 2012) or the principle of Contrast (Clark, 1993). These explanations do not account for the behavior in which familiar labels are paired with familiar objects, but only explain the novelty preference when a novel label-novel object pair is present.

None of the above-mentioned studies do, however, answer the question of very early word learning, which initiates much before the second year of life. Infants as young as 6 months are able to pair the most salient object in the scene to the most salient word in the stream they hear (Gogate & Bahrick, 2001; Shukla et al., 2011). From6 months on, they also recognize some of the most frequent words from their environment and accurately pair them with the visual stimuli (Bergelson & Swingley, 2012). The logical consequence of these results is to hypothesize that infants must posses the means to select the correct referents to the words they hear and that this process is fast and automatic. Our results confirm this hypothesis, showing that infants as young as 4 months quickly recognize the repeated label and pair it to the repeated object, but when the novel label occurs, they switch their attention to the novel object. It remains an open question whether this behavior is a consequence of the domain-general disjunctive syllogism or the possibly domain-specific mutual exclusivity. The exhibited behavior is, however, a complex response to the presented stimuli; it must involve a certain amount of decision-making based on the auditory input. We therefore suggest that the response is not only perceptually driven, but must be coordinated by a more complex cognitive process.

There is another possible mechanism for learning words, recently proposed under the name of cross-situational statistical learning (Smith & Yu, 2008). Under this frame, infants hold in mind the possible referential hypotheses and exclude the incorrect ones based the cross-situational statistics. The frame has received criticism regarding the number of hypotheses participants usually test (Trueswell et al., 2013), and regarding the true nature of the underlying mechanisms that guide cross-situational learning (Ichinco, Frank, & Saxe, 2009). In our experiment, infants have seen each object paired only with two other objects. Thus the probability of co-occurrence was 0.5, and the occurrence of all objects was matched for frequency. There is no relevant information about the possible label-referent pairs that could be extracted either cross-situationally or statistically. We can therefore conclude that cross-situational statistics could not contribute to the observed results in our experiment.

Just as previously published studies, Experiment 3 has tested the ability to label objects. Learning words for objects seem to in general precede learning words for other categories (Golinkoff & Hirsh-Pasek, 2008; Pruden et al., 2006). However, if another category is salient enough, 11-month-old infants seem to be able to associate the labels to that category too (Waxman & Booth, 2003). We therefore prepared a control experiment in which we question whether infants can instantaneously map labels also to movements.

2.3.2 Experiment 4

Labeling movements

Experiments 1 and 2 show that infants possess separate representational systems for objects and movements. Although infants' early vocabulary is predominantly consisted of nouns, i.e. words for objects, there are also some indications that infants are equally prone to label some other visual category, if it is salient enough (Waxman & Booth, 2003). We therefore hypothesized that the movements could also be labeled if they were the only salient category present in the experiment. We created a control experiment in which the only difference from Experiment 3 was that the category that is being labeled is movements.

2.3.2.1 Methods

2.3.2.1.1 Participants

In Experiment 4 we tested 38 infants. The final analysis contains the data of 24 infants who finished the experiment (11 boys, 13 girls, mean age 133 days, range 118 to 150). 14 infants were rejected from the analysis because they did not complete the task (n=10), had a strong bias to look to one side of the screen (n=3), or because of the eye-tracker failure (n=1). The criterion for the exclusion based on the side-bias was if the participant spent more than 75% of the time looking on one side of the screen. Other characteristics of subject sample are the same as in Experiment 1.

2.3.2.1.2 Materials

The auditory and visual materials were the same as those in Experiment 3. The only difference consisted in the fact that the labels from each word-list (Lists A and B, Table 1) were randomly assigned to 8 movements instead of objects. The relationship between a movement and a label was systematically maintained throughout the experiment.

2.3.2.1.3 Procedure

The procedure was equal to that of Experiment 3. The only difference was that during the test phase, infants saw two new objects, one moving in the same direction as the object in the familiarization phase, and the other moving in a new direction.

2.3.2.2 Results

In the test phase, infants on average started 69.6% (SD=20.4) of the trials by looking either to the center of the screen or away from the screen (paired two-tailed t-test against chance level (50%): t(23)=4.479, p=0.0001), as in Experiment 3. Again, we controlled for the first side-fixations during the test, and found no significant preference for either familiar or novel movements (ME(d)=-0.022, SD=0.191; paired two-tailed t-test against 0: t(23)=0.543, p=0.592, Cohen's *d*=0.157). As in Experiment 3, infants started to diverge their looking behavior after hearing the second label.

To assess the overall novelty preference, we computed the difference scores. The computation was the same as in Experiment 1. During the test phase of Condition 1, when they heard no word, infants showed no overall preference for either familiar or novel movements (ME(d)=-0.018, SD=0.389, paired two-tailed t-test against 0: t(23)=-0.225, p=0.824; Cohen's *d*=0.115). However, the looking pattern diverged after hearing the second word in Condition 2 and 3. In Condition 2, after they heard the repeated word, infants on average spent more time looking at the familiar movement (ME(d)=-0.159, SD=0.363, paired two-tailed t-test against 0: t(23)=2.136, p=0.043; Cohen's *d*=0.617). And after hearing the novel word in Condition 3, they looked significantly more at the novel movement (ME(d)=0.126, SD=0.285, paired two-tailed t-test against 0: t(23)=2.161, p=0.041; Cohen's *d*=0.624). As in Experiment 3, the analysis of variance between the three conditions revealed significant differences in looking behavior (one-way ANOVA with the factor Condition: F(2,69)=3.943, p=0.024). Post-hoc multiple comparison in Matlab showed that there were significant differences in looking behavior between Conditions 2 and 3 (familiar vs. novel word), but no significant differences between Conditions 2 and 3 and the Condition 1 (no word). The results of total looking time during the test phase in each of the conditions are presented in Figure 10.

Infants heard two different sets of words (12 infants per set) also in this experiment, but there is no significant difference between the two subgroups of subjects (two-way ANOVA with factors Condition and List showed the expected significant effect of Condition ($F(2,66)=4.018$, $p=0.027$) or List ($F(1,66)=2.157$, $p=0.147$), with no interaction between the factors ($F(2,66)=1.079$, $p=0.346$)).

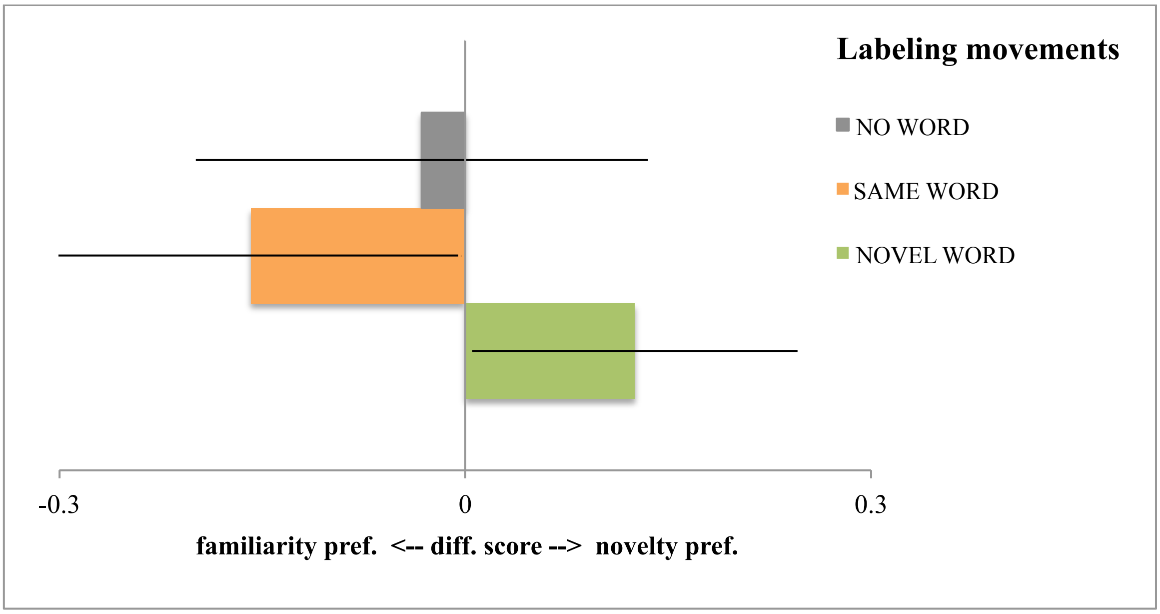

Figure 10 The results of total looking time after hearing the second word until the end of the test phase of Conditions 2 and 3 (and the matching time-window in Condition 1) in Experiment 4. The error bars represent 95% confidence intervals.

During the course of the test phase, infants on average switched their gazes between the two simultaneous videos 2.168-times (SD=0.452). The number of the switches between the two sides of the screen did not differ across conditions (the analysis of variance with the factor Condition: $F(2,69)=0.349$, $p=0.707$). As in Experiment 3, majority of the infants switched their gazes to the familiar video if their first fixation happened to be on the novel video in Condition 2, when they heard the familiar label (ME=75.35%, SD=26.65, paired two-tailed t-test against 50% chance level: $t(23)=4.659$, $p=0.0001$); similarly, infants switched their gazes to the novel video when hearing the novel word in Condition 3 if their first fixation was on the familiar video (ME=77.85%, SD=22.37, paired two-tailed t-test again 50% chance level: $t(23)=6.097$, $p= 3.214E-06$).

As in Experiment 3, we analyzed the looking behavior of infants during the test phase by a sample-based analysis of looks. The lines in Figure 11 represent the average looking behavior in all three conditions across the test phase. Infants' looking behavior started to diverge only ca. 1000ms after hearing the word: when hearing the repeated word, they turned their gaze more to the familiar movement (Condition 2), and when hearing the novel word, they switched their gazes to the novel movement (Condition 3). The response to the repetition of movements is overall relatively slower compared to the responses in Experiment 3 when the labels referred to objects. As in Experiment 3, we performed the permutation test between the looking behavior in Conditions 2 and 3 to identify the points in time during the test phase when infants' behavior differed significantly across different conditions (again we excluded Condition 1 because the overall behavior in Condition 1 did not significantly differ from 0 or from Conditions 2 and 3). Again, we used a standard permutation-based t-test with 10,000 permutations to calculate the mean and standard deviation of the overlap. We assumed a normal distribution, with the p-values lower than 0.05 called significant. The results of the permutation test are represented in the background picture of Figure 11, which shows the data-points in which the results from Conditions 2 and 3 differ significantly. Grey shades represent the significance levels of the differences between the two conditions.

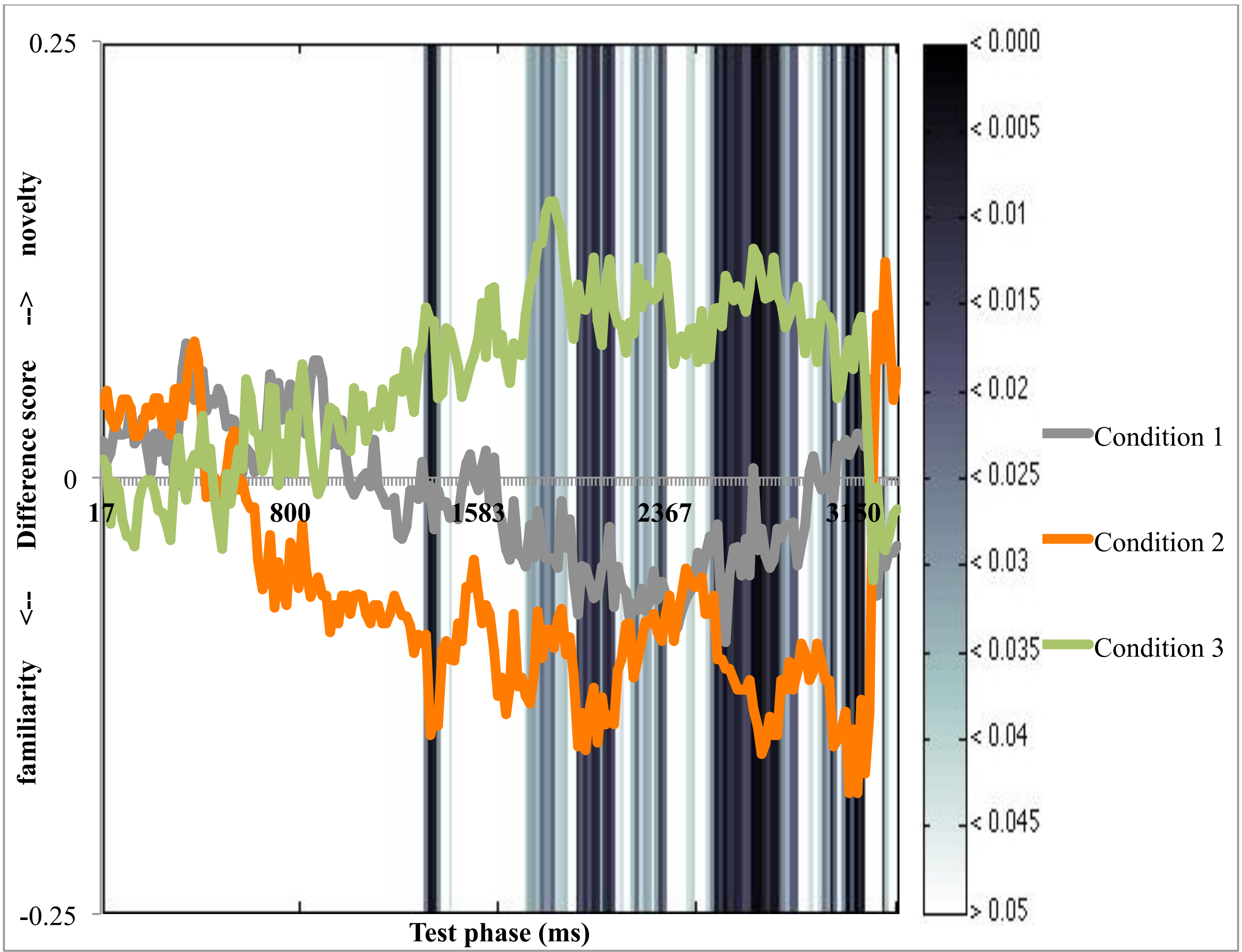

Figure 11 Sample based analysis of looking behavior in each of the three conditions during the test phase in Experiment 4. Positive scores signify looking at the novel movement, and negative scores looking at the repeated movement. Background figure represents the results of the permutation test between looking patterns in Condition 2 and 3. Grey shades represent the significance levels of the differences between the two conditions (as scaled in the color bar). The offset of the word in Conditions 2 and 3 happened at 800ms in each trial.

Because the stimuli in Experiment 3 and 4 differ only as to whether either objects or movements were labeled, we compared looking behavior during the test phase in both groups of infants in all three conditions. Two-way ANOVA with the factors Group and Condition revealed the expected significant effect of Condition ($F(2,138)=9.224$, $p=0.00017$), but no effect of the Group ($F(1,138)=0.284$, $p=0.595$), with no interaction between the groups ($F(2,138)=0.097$, $p=0.908$).

2.3.2.3 Discussion

As in Experiment 3, infants in this experiment showed no significant preference for either event on the screen when no word was repeated (Condition 1). They look longer to the repeated movement after hearing the repeated word (Condition 2), and look longer to the novel movement after hearing the novel word (Condition 3). We therefore conclude that the labeling mechanism that infants appear to use in Experiment 3 are operating also when infants discover that it is not the objects that are being labeled, but the movements. Labeling movements appears to be rapid and automatic, just as it is for objects.

There is however a small but important difference between the results in the two experiments: when objects are repeated, infants seem to detect the objects in their visual periphery and direct their fist gaze to them before hearing the second label, but when movements are repeated, this recognition is slower; thus looking patterns for repeated and novel words start to diverge later in time. This apparent delay has a very logical reason: temporal resolution is one of the key features of any movement or action – time is needed for a movement to happen, and therefore also for a movements to be recognized.

When we talk about labeling objects and movements/actions, the lexical categories with which they are commonly associated are nouns and verbs, respectively. In our experiments, movements are simplified linear movements to different directions in a 2D space. One could argue that verbal lexical distinctions for such linear movements are extremely unlikely in any natural language and that these experiments do not necessarily show any early verb learning ability. We agree with this argument to the extent that linear movements in different directions are most likely not lexicalized with verbs in natural languages. However, many natural languages do possess lexical items for the directions: in English, for example, the directions of movements are lexicalized with adverbs that accompany the verb (e.g. “up”, “down”, “left”, “right”). The present experiments were not designed to show learning of any specific lexical category, but to prove that names for objects are not the only entities that can be lexicalized in early infancy and that spatio-temporal dimension of any event can be represented independently and labeled independently from the objects that are part of the same event.

Infants in Experiment 1 (2.2.1) saw Object and Movement recognition trials in an interleaved manner, and they recognized the repetition in both. In Experiments 3 and 4 we separated the two conditions, so infants might have received each of the tasks as object-only or movement-only, forming a form of procedural memory for the task. Furthermore, infants hear labels for movements and objects in an interleaved manner in their everyday environment. An additional control would be needed to test whether infants would react to labeling in a similar manner also when labeling movements and labeling objects are mixed.

2.4

General discussion

The aim of Experiments 1 and 2 was to test whether infants encode and represent objects and the directions in which objects move as two separate entities, or as unified events in which movements help objects to be encoded and recognized (Baillargeon, 1987; Gredebäck & von Hofsten, 2007; Leslie et al., 1998). Our results show that infants can rapidly encode both objects and movements, and that the direction of movement is perceived as an independent feature of the visual scene. The representation of objects is stable after a short exposure and survives spatial transformations such as rotation. Thus movements are not perceived as a component of objects. This shows that infants did not recognize all repetitions as equally salient, but categorically responded to objects and movements. These results support the view that visual processing is segregated into two independent "what" and "how" pathways already during the first 4 months of life.

Experiments 3 and 4 were designed to address the issue of early labeling. As infants were recently shown to understand some words as early as 6 months after birth, we hypothesized they must possess the necessary cognitive abilities to recognize a labeling situation and map the most probable label-referent pairs. To be able to show that the segregated representation of objects and movements can be extended also to labeling situations, we used the same visual stimuli as in previous experiments and added bisyllabic labels. The results show that infants recognize object-label pairs after a single exposure. When the label is repeated, they look longer to the repeated object. However, when the label changes, they appear to logically infer that the new object on the screen is being labeled and look longer to it. Thus, looking responses of 4-month-olds to labeling are very rapid and yet very elaborated. Moreover, infants show similar responses also to situations in which movements are labeled, indicating that they have no overall preference for labeling objects, but that they are able to map a label to whatever the most plausible referent they find. Of course, these experiments do not answer the question of word learnability in general: nouns constitute the majority of infants' early vocabulary, and there is a variety of reasons for it (Golinkoff & Hirsh-Pasek, 2008). And although recent experiments have shown that infants know certain common words at the age of 6 months, the majority of infants' vocabulary still develops much later, during and after the second year of life. We therefore suggest that the ability to label is separated from other abilities that are necessary for word learning, such as understanding the social context of labeling (Csibra & Gergely, 2009; Csibra, 2010), the ability to categorize (Spelke & Kinzler, 2007), and the long-term retention of words' meaning (Feigenson & Halberda, 2008; Spiegel & Halberda, 2011).

A crucial part of labeling is the ability to map new labels to novel referents. There is still an on-going debate about whether this ability is governed by a general novelty preference, or by the principle of mutual exclusivity, just as it is unclear whether mutual exclusivity is an innate language-specific principle, or a more domain-general principle shared with other animals. Whereas our experiments might not address the exact nature of logical reasoning present in infants, we believe that the responses of infants in these experiments can be best explained by the use mutual exclusivity (see Ch. 2.3.1.3). Again, our experiments do not directly address the question whether such principle is innate or learned. They do however show that it is operational in infants as young as 4 months.

Infants' ability to map novel labels to novel objects in ambiguous situations has been until now shown in 12-moth-olds (Smith & Yu, 2008). How is it possible that so much younger infants in our study succeeded in the task? One possibility is that the present study has only focused on the process of labeling, but did not require any memorizing. Most previous studies (but see Spiegel & Halberda, 2011) test the understanding of words only after a certain period of time, when words are already stored in long-term memory. While preverbal infants show long-term memory for words since birth on, and even memory for the arbitrary label-object pairs (Bahrick & Pickens, 1995; Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012; Gogate & Bahrick, 2001), long-term retention of multiple label-referent pairs has not been tested. It is therefore possible that labeling is feasible in young infants, but memorizing multiple pairs becomes too demanding. And since long-term retention was not required in our tasks, infants were successfully labeling both familiar and novel items. Another possibility is that the present studies have only tested the recognition and labeling of previously unknown objects and labels. Infants thus recognized the objects, movements, and labels with which they were familiarized during the experiment, but they did not have any stable preceding knowledge about any of these items. When much older infants were tested with familiar and novel label-object pairs (Halberda, 2003), such conceptual difference might have impeded them from correctly pairing the novel labels to the novel referents.

There are of course additional questions that remain open after Experiments 3 and 4. First, the development of logical thinking has been most commonly associated with the development of a symbolic representational system, i.e. language (Spelke, 2003). Recently however, 12-month-old infants were shown to be able to reason about possible and impossible events even when no linguistic stimuli were present (Cesana-Arlotti et al., 2012; Téglás et al., 2011, 2007). It is unclear whether infants would respond in the same way if non-linguistic auditory stimuli were associated with the visual input presented in the above experiment. Second, infants in our experiments perceived full repetitions of auditory and visual stimuli (each word was invariantly repeated through the experiment). Although infants were shown to be able to categorize both auditory and visual input at a very early age (Ferry et al., 2013; French, Mareschal, Mermillod, & Quinn, 2004; Sebastián-Gallés, 2006), it is unclear whether and how the increased complexity of the stimuli would affect the labeling situation. Third, most word learning happens in communicative situations, with the help of ostensive cues, which are shown to facilitate infants' learning (Csibra, 2010). However, there seem to exist certain situations in which infants are misled by ostensive cues (Topál, Gergely, Miklósi, Erdohegyi, & Csibra, 2008). It is therefore unclear how the labeling situation presented in our experiment could be affected by additional ostensive cues, i.e. whether the labeling situation would improve or worsen if more ostensive cues were added.

We can nonetheless conclude that infants at the age of 4-5 months are able to label both abstract objects and directions of movement. The results of the experiments described in this chapter also indicate that infants process their visual scenes rapidly and that they reason logically about possible label-referent pairs. However, these experiments did not address the questions of a) how can infants memorize multiple label-referent pairs for a longer period of time, and b) what are the possible processing limitations that infants could have in these tasks. In the next chapter, we will contrast the word-learning task with the learning of auditory regularities in order to assess possible differences in learning.

3

Word-learning and other cross-modal learning processes

3.1 Introduction

It was recently shown that newborns' retention for the words presented auditorily lasts at least some minutes after the familiarization, although the processes of storing the memories and recalling can be interfered by other sounds (Benavides-Varela et al., 2011, 2012). During the first year, the ability to retain words progressively extends to longer periods of time (Jusczyk & Hohne, 1997), and so does the ability to retain familiar visual events, such as objects and motions (Bahrick & Pickens, 1995; Fagan III, 1973). By the age of 6 to 7 months, infants already show sensitivity to arbitrary syllable-object pairs and retain them for a longer period of time (Gogate & Bahrick, 2001; Shukla et al., 2011). It is therefore not so surprising that 6-9-month-old infants were recently shown to recognize common labels for the objects in their environment (Bergelson & Swingley, 2012). While there has been a large body of work done on the role of saliency of both referents and labels in early word learning when one label-object pair is present (Brent & Siskind, 2001; Pruden et al., 2006), less is known about learning multiple words in noisy conditions, such as natural language learning environment is.

Intermodal learning has received attention also from another perspective. McMurray and Aslin (McMurray & Aslin, 2004) proposed that infants' ability to learn about various visual or auditory categories could be assessed more easily by a paradigm that would avoid the reliance on one response only, which is the case in the habituation paradigms. Instead, they propose an eye-tracking paradigm that stimulates infants to anticipate the visual reinforcer each time a member of the learned visual or auditory category appears. Several studies followed in which it was shown that infants can associate a learned category predominantly from the auditory domain with a specific position (on the screen), where the visual reinforcer could appear (Albareda-Castellot, Pons, & Sebastián-Gallés, 2010; Benavides-Varela et al., 2012; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011; Kovács & Mehler, 2009b). However, whenever two regularities were presented in an interleaved manner, each associated with a different position of the reinforcer's appearance, (monolingual) infants have predominantly shown to have learned one regularity only, whereas they did not exhibit any specific behavior when the other regularity was presented (Kovács & Mehler, 2009b). One of the proposed explanations is that infants find it difficult to switch between two structures because of the cognitive control needed during this process – they have no problem learning how to associate multiple auditory regularities to visual reinforcers when they are presented consecutively (Kovács & Mehler, 2009a). It is also possible or that they find different structures too similar to each other, therefore mixing the two (Kovács & Mehler, 2009b). The fact that infants are able to learn multiple regularities when they are presented only auditorily (Gerken, Balcomb, & Minton, 2011) points also to the possibility that infants find cross-modal learning of multiple regularities too demanding.

Preverbal infants learn the first labels in a relatively noisy natural environment, where they hear various words that label various objects and actions in a short amount of time, almost simultaneously (Bergelson & Swingley, 2012). 7-month-old infants were also shown to be able to learn more than one label-object pair in a habituation procedure (Gogate, Bolzani, & Betancourt, 2006). It is therefore possible that two different processes underlie learning multiple labels and learning multiple auditory regularities, even if both tasks are intermodal. To explore this hypothesis in more detail, we created 3 eye-tracking experiments for 8-month-old infants. The aim of the first two experiments was to disentangle possible reasons for infants' inability to learn two regularities simultaneously, such as the nature of the regularities and possible interferences between them. The aim of the third experiment was to test how fast do infants learn two labels simultaneously. All three experiments involve learning to anticipate a visual outcome after hearing an auditory stimulus: in the first two, the reinforcers appear only after the auditory stimuli, and in the third one, the referents move only after being labeled. In order to assess possible differences, we compared looking patterns in the word-learning and in the regularity-learning experiments.

3.2

Learning auditory regularities at 8 months

3.2.1 Experiment 5

Learning about lexical stress regularities

The present experiment asked whether infants learn two distinct stress patterns simultaneously. Lexical stress patterns are regularities that are present in many languages. Infants start being sensitive to lexical stress regularities in the second half of their first year of life (Friederici, Friedrich, & Christophe, 2007; Jusczyk, Cutler, & Redanz, 1993; Skoruppa, Cristià, Peperkamp, & Seidl, 2011). This sensitivity is shown to help infants in word segmentation and word learning processes (Gerken & Bollt, 2008; Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999). It is also a regularity that does not rely on repetition patterns, unlike other frequently studied regularities (Johnson et al., 2009; Marcus, Vijayan, Bandi Rao, & Vishton, 1999), but is simply a positional regularity. Our hypothesis was that infants might learn multiple regularities that are more closely related to language more easily than more abstract repetition-based regularities. If they learn two auditory regularities as well as their association to the visual reinforcers, they might anticipate the appearance of the reinforcers correctly for both regularities. To test this hypothesis, we used an adapted version the anticipatory eye movements procedure, introduced by McMurray and Aslin (McMurray & Aslin, 2004) and developed by Kovacs (Kovács, 2008), in which two auditory regularities are associated with two sides of the screen where the reinforcers can appear.

3.2.1.1 Methods

3.2.1.1.1 Participants

In Experiment 5 we tested 40 infants. The final analysis contains the looking-behavior of 20 infants who finished the experiment (11 boys, 9 girls, ME=239 days, range 231 to 257, SD=11.4 days). 20 infants were rejected from the analysis because they did not complete the task (n=11), had a strong bias to look to one side of the screen (n=5), or due to the eye-tracking failure (n=4). The criterion for the exclusion based on the side-bias was if the participant spent more than 75% of the time looking on one side of the screen. All infants were born without complications and none of the parents reported their infant having visual or auditory problems. The APGAR values for all infants were 8 and above. Informed consent was obtained from the parents of all infant participants before the beginning of the experiments. The infant experiments were approved by the Bioethics Committee of SISSA – International School for Advanced Studies (date of the approval 25/03/2010). Parents received a small monetary compensation for travelling expenses.

3.2.1.1.2 Materials

The auditory stimuli consisted of 8 word pairs that differed only in lexical stress position. There were thus totally 16 words; 12 words were used in familiarization and 4 in the test phase. All the words were nonce, bi-syllabic; all syllables had consonant–vowel (CV) structure. Stimuli were recorded by a female native speaker of Italian and normalized for duration and intensity using the Praat speech analysis software (final word length for stress-initial words was 610ms, final word length for stress-initial words was 460ms; output intensity 70dB).

Visual stimuli that were used as reinforcers consisted of 8 looming pictures of puppets on a white surface; 6 of them were presented during familiarization and 2 during the test phase. The initial size of the picture was 2.0º of visual angles and it was loomed till the size of 10.0º. The reinforcer pictures were randomly presented at one of the two sides of the screen. For each stress-position, the side of the appearance of the reinforcers was fixed within subjects. In total, infants heard each word and saw each of the reinforcers 3 times. The stimuli are presented in Table 2.

STRESS-INITIAL WORDS	STRESS-FINAL WORDS
‘BERI	BE’RI
‘CINO	CI’NO
‘GUBA	GU’BA
‘MOCE	MO’CE
‘SOVA	SO’VA
‘PULE	PU’LE
‘SIPO	SI’PO
‘TANI	TA’NI

Table 2 The stimuli used in the Experiment 5.

3.2.1.1.3 Procedure

In each trial, infants saw two small white windows (2.0º) on a grey background. After hearing either a stress-final or a stress-initial word, there was a fixation period of maximum 3500ms. If they fixated the white window at the correct side of the screen, the reinforcer appeared and started to loom from the position of one of the two white windows. If they did not fixate the correct window during this period, no reinforcer appeared and the next trial started (see Figure 12). Gaze-contingent reinforcers were introduced because this enabled us to have a procedure more similar to the conditioned head-turn procedure which is frequently used in tasks that involve learning auditory regularities (Kuhl, 1991). In head-turn procedure, when two auditory regularities are learned simultaneously (as in Minton et al 2011), the third one has to be introduced in the test in order to elicit novelty preference. The advantage of this paradigm is that instead of eliciting novelty preference, infants are taught to connect auditory and visual input and therefore actively anticipate the outcome. There were 36 familiarization trials and 8 test trials; the stress-initial and stress-final trials were presented interleaved, with no more than two exemplars of each condition presented consecutively. The conditions were randomly assigned to the side of the screen where the reinforcer would appear. Thus half of the infants received reinforcers on the left side after hearing stress-initial words and on the right side after hearing stress-final words. The other half of the infants received the words from each regularity on the opposite sides. Familiarization and test trials were equal in structure, except that the new set of visual and auditory stimuli were introduced in the last 8 trials. The trials were separated by central fixations and the experimenter initiated each trial when infants' gaze was directed to the center of the screen. Infants' gaze was recorded with a TOBII T60 eye-tracker (http://www.tobii.com/) at a rate of 60 Hz. The eye-tracker was integrated into a 17-inch TFT screen. The calibration procedure was the same as described in Experiment 1. The stimuli were presented via EventIDE software (http://okazolab.com). Gaze recordings were segmented and grouped using Perl program; statistical analysis was run with Matlab 7.9.0 software.

Infants were seated on their parent's lap (in order to avoid any discomfort due to a new environment without a presence of a familiar person) at about 60cm distance from the monitor. Parents wore blocked glasses to avoid the eye-tracker collecting their gaze and to ensure that infants' looking-behavior was not affected by parental influence. To determine whether infants' looking behavior to the two test videos in each trial differed, we determined a region of interest that matched the size and the location of the videos on the screen (480x480 pixels invisible square regions on each side of the screen). Only the looks that fell into these ROIs were counted in the measures of fixations. The threshold for fixation was 200ms of uninterrupted looking to the same area; we measured the first looks and anticipatory orientation latencies: the first fixations to the correct ROI (which consequentially triggered the reinforcer) and the time needed to reach the correct side of the screen.

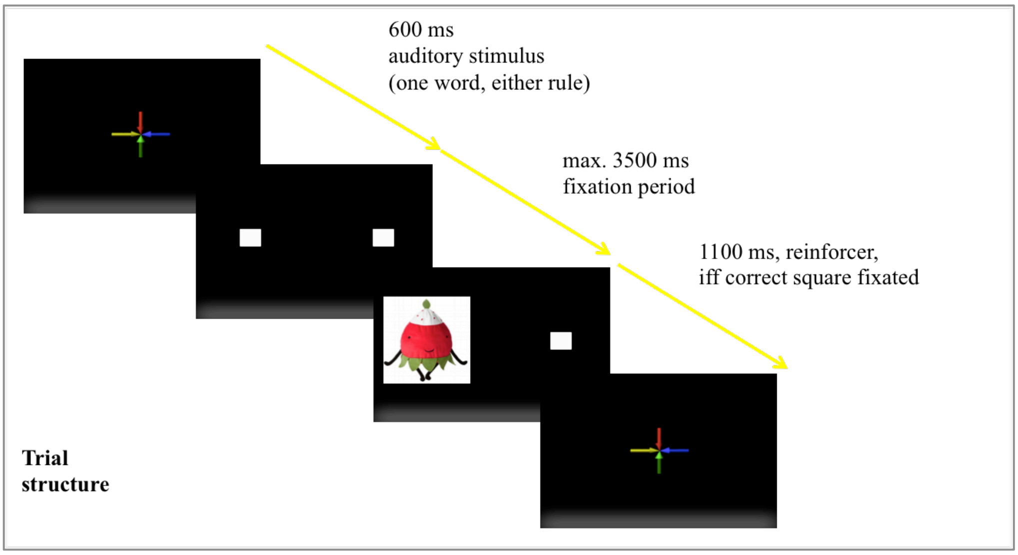

Figure 12 The structure of the trial in Experiment 5.

3.2.1.2 Results

The results show that infants successfully learned to associate one of the two regularities with their visual reinforcers, similarly to the experiments where repetition-based regularities were taught (Kovács, 2008). The analysis of the first fixations in the test phase shows that infants directed their anticipatory first looks significantly more towards the correct side of the screen then to the incorrect one when they heard the words that belonged to the stress-final word category (difference score was computed as described in Experiment 1, with 0 meaning no preference: ME(d)=0.38, SD=0.69; paired two-tailed t-test against 0: t(19)=2.407, p=0.0267; Cohen's *d*=0.760; Figure 13). To answer the question whether the proportion of correct anticipatory first looks changed during the course of the experiment, we performed linear correlation analysis. In the stress-final condition, the proportion of correct first looks significantly increased during the experiment (correlation coefficient (corrval)=0.49, p=0.021).

When they heard the words that belonged to the stress-initial word category, the first looks were, however, not significantly directed towards the correct side (ME(d)=0.233, SD=0.712; paired two-tailed t-test against 0: t(19)=1.465, p=0.159; Cohen's *d*=0.463) (see Figure 13). The proportion of correct first looks in the stress-initial condition did not change significantly across trials (corrval=-0.101, p=0.652) (see Figure 14). Subject-based analysis revealed that 5 out of 20 participants learned both regularities, 6 exhibited learning only in the stress-final condition, and 5 in the stress-initial condition, whereas the remaining 4 were at chance.

However, the difference in the looking behavior across conditions was, overall, not very big. Although infants responded significantly correctly when hearing stress-final words, but not when hearing stress-initial words, the difference between the conditions was far from significant (two-sample t-test between the conditions: t(38)=0.665, p=0.51). The average response latency was 950ms (SD=160) and there is no significant difference in the response latencies between the two conditions (two sample t-test: t(38)=0.995, p=0.331). There was also no significant decrease or increase of response latencies across trials in none of the conditions (stress-final condition: corrval=0.213, p=0.341; stress-initial condition: corrval=0.394, p=0.0695; see Figure 15). Infants did not change significantly their looking behavior across trials in the stress-final condition: on average 64,4% (SD=7.9) of infants were actively searching for the reinforcer at either of the ROIs. In the stress-initial condition, the proportion of infants who actively searched for the reinforcer was 77.5% on average (SD=9.4) and decreased significantly towards the end of the experiment (corrval=-0.571, p=0.005).

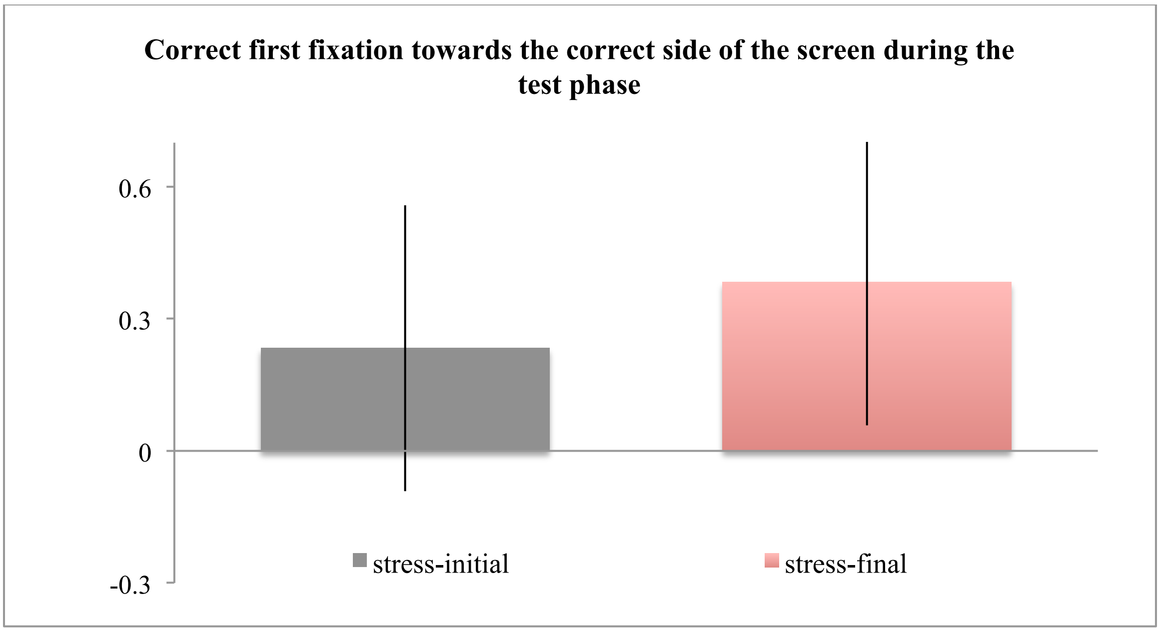

Figure 13 First fixations towards the correct side of the screen during the test phase in each condition. The error bars represent 95% confidence intervals.

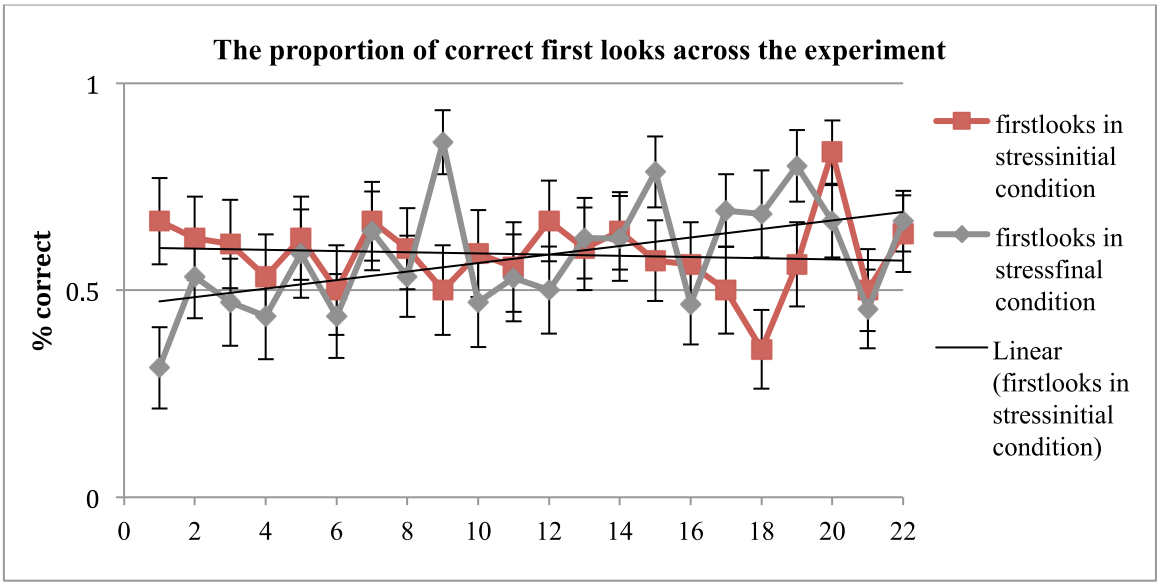

Figure 14 The proportion of first looks to the correct side of the screen across the trials in each condition. The last 4 trials in each condition are the test trials. The error bars represent the standard errors.

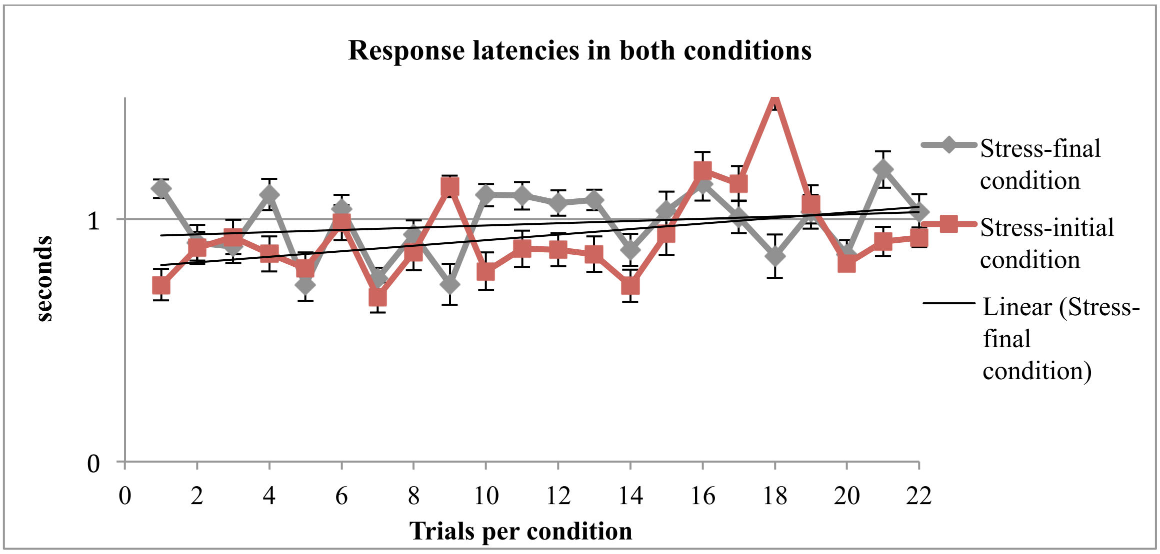

Figure 15 The average time infants needed to fixate the correct side of the screen, which triggered the appearance of the reinforcers, in each condition. The error bars represent standard errors.

3.2.1.3 Discussion

The results of Experiment 5 show that infants have learned the association between the side of the screen where the reinforcer appears and the auditory regularity – but only for one of the two regularities presented. Although the lexical stress position is a language-based regularity that is salient to infants in the second half of the first year, and we hypothesized will help infants to learn the two regularities simultaneously, they only exhibited the predicted behavior after hearing stress-final words. This result is consistent with the findings in other studies where infants were presented with multiple auditory regularities and received visual reinforcers (Kovács & Mehler, 2009b).

Why did infants learn only stress-final, and not only stress-initial words? The majority of the infants we tested came from Italian speaking environment, and the vast majority of bisyllabic words that infants hear in Italian is stressed on the initial syllable. In the Italian corpus that we use in the word segmentation study presented in Chapter 4 of this thesis, only 5.35% of all bisyllabic words were stressed on the final syllable. This follows from the fact that lexical stress in Italian is predominantly on penultimate syllables (Den Os & Kager, 1986). Italian infants are therefore most probably more familiar with stress-initial words. Their attention directed towards stress-final words can be therefore the consequence of infants' general attention to novel stimuli (e.g. Marcus, Vijayan, Bandi Rao, & Vishton, 1999). This hypothesis is confirmed by the fact that the proportion of infants who actively participated in each trial decreased only during the stress-initial condition: they may have habituated to the more familiar stimuli faster than to the non-familiar stimuli.

However, the fact that there was no significant difference in the responses between the two conditions could indicate that they have the tendency to learn both, just as predicted in our initial hypothesis. This experiment does, thus, not answer the question about what are the possible factors that contribute to the fact that infants seem to have difficulties with responding accurately to multiple regularities simultaneously. It is still unclear whether the results are due to the developmental stage in which the cognitive control needed to switch between two regularities (and to decide about the location of the visual reinforcer) is still developing. Our result could also be the consequence of infants' inability to discriminate the stimuli well enough to avoid interferences between the categories. To control for this possibility, we have designed an experiment in which the linguistic regularity was contrasted with an auditory regularity composed of non-linguistic stimuli. Infants' brain responses to linguistic and non-linguistic stimuli are very different (Benavides-Varela et al., 2011; Dehaene-Lambertz et al., 2010; Peña et al., 2003), and infants respond to non-linguistic stimuli categorically (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). We therefore hypothesized that introducing non-linguistic stimuli might facilitate learning multiple associations between auditory and visual stimuli.

3.2.2 Experiment 6

Learning about a linguistic and a non-linguistic regularity at 8 months

The non-linguistic stimuli used in this experiment were different bird songs. They were used because we wanted to contrast two codes of communication that are perceived by the infants of its species in a specialized way: speech has a special role in perception for human infants as well as birdsongs for the bird infants, and there are many structural parallels between the two codes (Doupe & Kuhl, 1999; Mehler, Nespor, & Peña, 2008). They were also used because infants fine-tune their listening to speech by the age of 6 months and react to other-species' sounds in a categorically different way (Ferry et al., 2013).

3.2.2.1 Methods

3.2.2.1.1 Participants

In Experiment 8 we tested 30 infants. The final analysis contains the looking-behavior of 20 infants who finished the experiment (7 boys, 13 girls, ME=247 days, range 229 to 269, SD=9,69 days). 10 infants were rejected from the analysis because they did not complete the task (n=7) or had a bias to look to one side of the screen (the criterion for the exclusion was more than 75% of the total time spent only on one side of the screen) (n=3). Other characteristics of the subject sample are the same as in Experiment 5.

3.2.2.1.2 Materials

Half of the visual and auditory stimuli were equal to the stimuli in Experiment 5: for the linguistic regularity, stress-final words from Experiment 5 were used, and the reinforcers associated to them were the 8 puppets presented in the previous experiment. The other half of the stimuli consisted of 8 different bird-songs, normalized for duration and intensity using the Praat speech analysis software (d=1000ms, output intensity 70dB). The reinforcers associated to the non-linguistic regularity were 8 different bird images of the same size as the images of the puppets, presented on a white surface. As in the previous experiment, once the reinforcer was triggered, the image loomed from 2.0º to 10.0º and stayed visible for the total of 1100ms.

3.2.2.1.3 Procedure

The procedure was equal to the procedure in the Experiment 5, except that the number of familiarization trials was decreased to 24, because of the higher rejection rates due to the incompletion of the experiment in the previous experiment. Thus, each stimulus was repeated only twice during the familiarization phase.

3.2.2.2 Results

The analysis of the fist fixations in the test phase shows that infants directed their anticipatory first looks significantly more towards the correct side of the screen only when they heard the words (difference score was computed as described in Experiment 1, with 0 meaning no preference: ME(d)=0.292, SD=0.527; paired two-tailed t-test against 0: t(19)=0.203, p=0.04, Cohen's *d*=0.783). When they heard the sounds that belonged to non-linguistic category, the first looks were not significantly directed to any side of the screen (ME(d)=-0.188, SD=0.456, t(19)=1.425, p=0.171, Cohen's *d*=0.582) (see Figure 16). Subject-based analysis revealed that 4 out of 20 participants learned both regularities, 7 responded with the first look to the correct side only when hearing words, whereas only one participant responded only to the bird songs. The remaining 8 were at chance.

Contrary to the results in Experiment 5, infants in this experiment responded significantly different to the linguistic and non-linguistic stimuli. The results in the test phase differ significantly in the two conditions (two-sample t-test: t(38)=2.567, p=0.014). We also measured the time needed for the gaze to fixate the correct part of the screen (response latency). The average response latency was 890ms (SD=278), but infants' responses were significantly faster in the word condition (ME=787ms, SD=232) than in the bird-song condition (ME=994ms, SD=282; t(15)=2.684, p=0.017; Figure 17). However, there was no significant decrease or increase of response latencies across trials in any of the conditions (corrval<0.08, p>0.63). We also measured whether the proportion of correct anticipatory fist looks changed during the course of experiment. The proportion of correct first looks did not change significantly in none of the conditions (Word condition: corrval=0.42, p=0.065; Bird-song condition: corrval=-0.294, p=0.269). Infants did not change significantly their looking behavior across trials in the stress-final condition: on average 91.9% (SD=5.5) of the infants were actively searching for the reinforcer at either of the ROIs. While the proportion of infants did not decrease in the bird-song condition, significantly fewer infants were actively participating in the words condition towards the end of the experiment (corrval=-0.516, p=0.04).

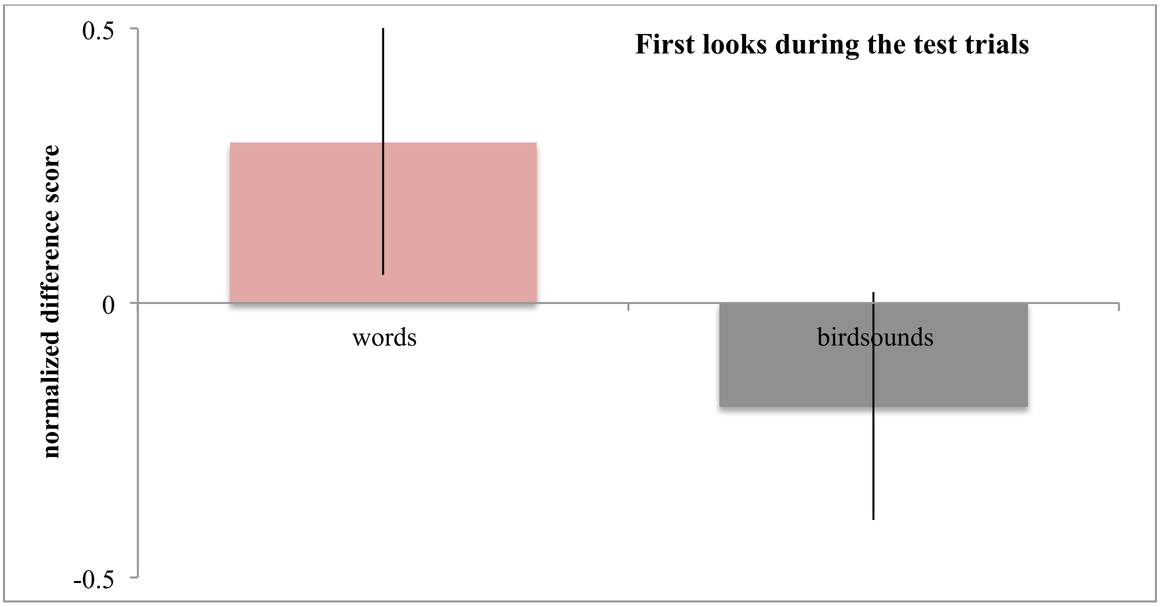

Figure 16 The average difference between first looks to the correct side and the incorrect side during the test trials in both conditions. The error bars represent 95% confidence intervals.

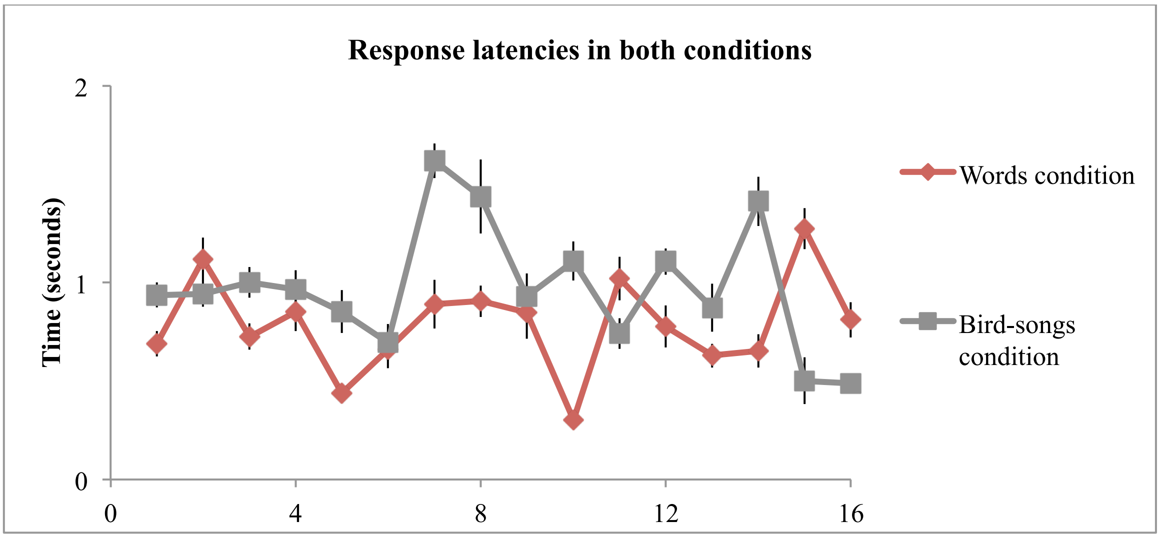

Figure 17 The average time infants needed to fixate the correct side of the screen, which triggered the appearance of the reinforcers, in each condition. The error bars represent standard errors.

3.2.2.3 Discussion

Infants in this experiment showed a strong overall preference for the linguistic stimuli over the non-linguistic sounds, such as bird-songs. Their first fixations were directed towards the correct side of the screen whenever they heard words; even when hearing bird-songs, their first gaze tended to go to the side of the screen where the reinforcer for the words appeared. The preference was persistent although their attention for the word stimuli (but not for the bird-song stimuli) slightly dropped towards the end of the experiment. Their responses were overall faster when hearing the words than when hearing the bird-songs. These results confirm previous results that show infants' preference for speech over non-speech stimuli (Ferry et al., 2013; Vouloumanos & Werker, 2004, 2007).

The results from Experiment 5 and 6, taken together, indicate that infants cannot learn to associate multiple auditory regularities to multiple visual reinforcers, and that this inability cannot be explained by the nature of the input. Even if the differences are perceptually very salient, infants are not triggered to actively show learning of both regularities simultaneously. They are not compelled to show learning, even if the appearance of visual stimuli is dynamically presented only when infants actively participate in the experiment (gaze-contingent conditioning). We therefore conclude that infants lack the cognitive control needed to successfully switch between two regularities.

This brings us back to the question of word learning. If infants learn multiple words during the second half of the first year of life in the relatively noisy natural language environment, do they learn multiple words during the same event, or do they learn them consecutively, one by one? Do they exhibit similar cognitive control limitations as when learning multiple regularities? To answer these questions, we tested another group of infants with multiple label-object pairs.

3.3

Learning words at 8 months

3.3.1 Experiment 7

Learning about label-object pairs

In the present experiment we addressed the question of learning multiple label-object pairs simultaneously. A similar experiment was recently conducted with 12-month-old infants, showing that infants are able to derive which label refers to which object after a brief exposure (Smith & Yu, 2008). The scope of the present experiment was to see a similar learning with younger infants. To eliminate the need of remembering multiple label-object pairs across many trials and to create an environment in which infants would create expectations and thus actively anticipate the events, we used conditional learning as in the previous two experiments.

3.3.1.1 Methods

3.3.1.1.1 Participants

In Experiment 7 we tested 33 infants. The final analysis contains the looking-behavior of 20 infants who finished the experiment (8 boys, 12 girls, ME=254 days, SD=8.18, range 239 to 271). 13 infants were rejected from the analysis because they did not complete the task (n=5) or had a strong bias to look to one side of the screen (n=8). Other characteristics of the subject sample are the same as in Experiment 5.

3.3.1.1.2 Materials

The auditory stimuli consisted of 6 words, two for each of the three blocks. All the words were nonce, bi-syllabic; all syllables had the consonant–vowel (CV) structure; combinations of syllables that are easily discriminable were chosen to facilitate learning. Stimuli were recorded by a female native speaker of Italian and normalized for duration and intensity using the Praat speech analysis software (final word length was 610ms; output intensity 70dB).

Visual stimuli were 6 animated puppets on a white surface, 2 were used for each block. The blocks were randomized across participants. The frame size was 10.0º for each puppet. In each trial, infants saw two puppets simultaneously, one at each side of the screen. Across trials, the position of the puppets was randomized so that each puppet appeared 5 times on each side. Infants initially saw still frames of the movies, while the animation started only if they fixated the correct puppet in each trial. The movements of each pair (during the animation) were matched for speed and tempo. Each puppet was presented 10 times, and each label was presented 5 times, 4 times in the familiarization phase and once in the test phase. The stimuli are presented in Table 3.

	LABEL	REFERENT
PAIR 1	'BIGO	VIOLIN CORN
	'DAKU	FOX
PAIR 2	'DAFE	LIZARD
	'PUNA	MONKEY
PAIR 3	'LIME	DUCK
	'RAZI	WEASEL

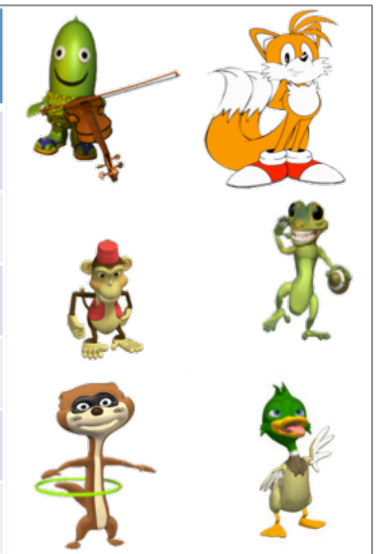

Table 3 The stimuli used in Experiment 7. The images are the still frames of the movies that infants saw during the reinforcement phase.

3.3.1.1.3 Procedure

Instead of one familiarization and one test phase we built 3 blocks, each containing 8 familiarization and 2 test trials. In each block, infants repeatedly saw same two puppets and heard one of the two words; in total, each puppet was labeled 5 times. In each trial, infants saw two still frames of the movies of the puppets. Across trials, the position of the puppets was randomized so that each puppet appeared 5 times on each side of the screen.

100ms after the onset of the visual stimuli, one of the two words was presented. After hearing the word, there was a fixation period of maximum 3500ms. During the familiarization trials, the puppet was animated for 2000ms if the infant fixated the correct puppet. If s/he did not fixate the correct puppet during this period, there was no animation and the next trial started. In most of the trials infants looked to both puppets, triggering the movement, which caused learning about which label is paired with which puppet. Gaze-contingent reinforcing was introduced because this enabled us to teach infants about the correct label-object association even if more than one possible referent was present in the visual scene. It also enabled the object-label associations that were not defined spatially because the side at which the puppet appeared was changing. During the test trials, the puppets remained still the whole time of the fixation period, which enabled us to see whether infants look at the correct object for each label. The trials were separated by central fixations and the experimenter initiated each trial when the infants' gaze was directed to the center of the screen.

Infants' gaze was recorded with a TOBII T60 eye-tracker (http://www.tobii.com/) at a rate of 60 Hz. The eye-tracker was integrated into a 17-inch TFT screen. The calibration procedure was the same as in previous experiments. The stimuli were presented via EventIDE software (http://okazolab.com). Gaze recordings were segmented and grouped using Perl program; statistical analysis was run with Matlab 7.9.0 software. Infants were seated on their parent's lap (in order to avoid any discomfort due to a new environment without the presence of a familiar person) at about 60cm distance from the monitor. Parents wore blocked glasses to avoid the eye-tracker collecting their gaze and to ensure that infants' looking-behavior was not affected by parental influence. To determine whether infants' looking behavior to the two test videos in each trial differed, we determined a region of interest that matched the size and the location of the videos on the screen (480x480 pixels square regions on each side of the screen). Only the looks that fell into these ROIs were counted in the measures of fixations. The threshold for fixation was 200ms; we measured the response latencies: the first fixations to any of the defined ROIs and the first fixations to the correct ROI (which consequentially triggered the reinforcer). During the test trials, we also measured the total looking time and the longest uninterrupted fixation to each of the two puppets.

3.3.1.2 Results

The results show that infants learned to associate two words to two objects. On average, first fixations during the test trials of the three blocks were directed significantly more towards the object that was labeled than towards the non-labeled one (difference score was computed as described in Experiment 1, with 0 meaning no preference: ME(d)=0.092, SD=0.155; paired two-tailed t-test against 0: t(19)=2.649, p=0.0158; Cohen's *d*=0.837). Because the objects did not move in the test phase, we could also measure the total looking time infants spent observing the objects. They looked overall longer to the correct object, but only marginally significantly (ME(d)=0.154, SD=0.358; paired two-tailed t-test against 0: t(19)=1.92, p=0.0699; Cohen's *d*=0.607). However, the longest uninterrupted fixation was directed significantly more often to the correct object (ME(d)=0.209, SD=0.365; paired two-tailed t-test against 0: t(19)=2.556, p=0.0193; Cohen's *d*=0.808) (Figure 18).

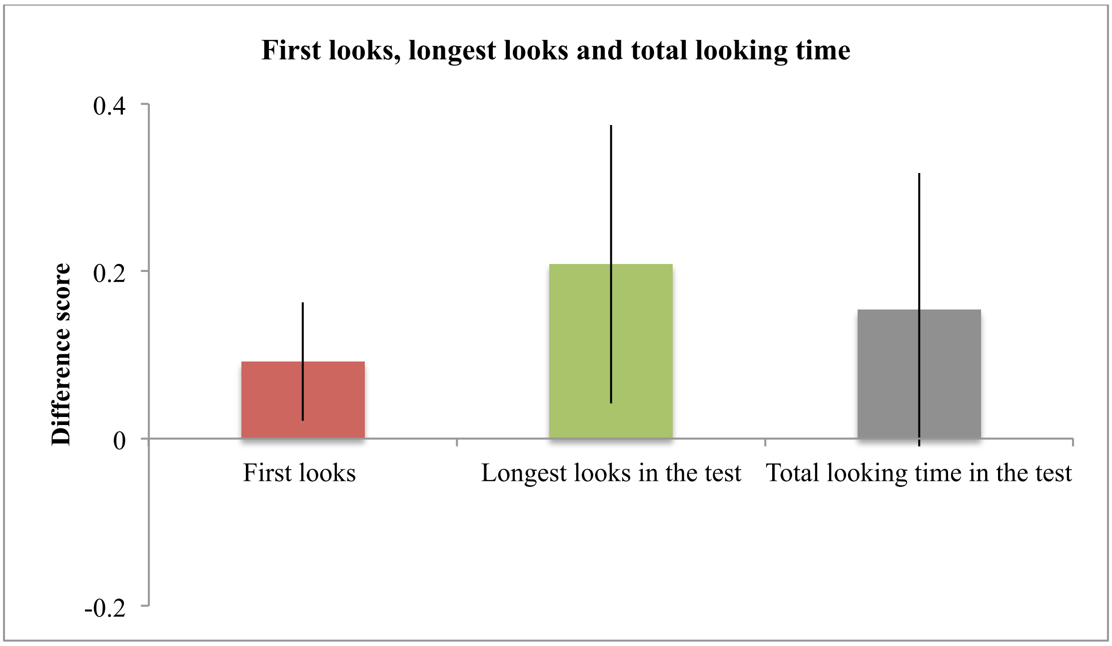

Figure 18 The first bar represent the overall difference between correct and incorrect first looks during the whole experiments and the difference between correct and incorrect longest uninterrupted fixations and overall looking time during the test phases. The error bars represent 95% confidence intervals.

The presented results include infants' responses to all 6 label-object pairs together. In order to verify whether infants responded to some label-object pairs more actively than to others, we performed one-way ANOVAs between (both correctly and incorrectly directed) total looking times and longest uninterrupted looks during the test phases for each of the presented objects. These revealed no differences between the objects ($F(5,84)<0.71$, $p>0.61$), indicating that all the presented label-object pairs were learned equally well, unlike the label-object pairs presented in Smith & Yu (Smith & Yu, 2008). Because the item-based analysis alone does not answer the question whether individual participants in general learned both labels during each block or not, we also verified whether the longest uninterrupted looks in both test trials in each block were predominantly oriented towards the correct object or not. Twelve out of twenty infants looked to both puppets correctly at least during one block: 9 in both test trials in one block, 2 infants in two blocks, and 1 infant in all three blocks.

3.3.1.3 Discussion

The results of this experiment show that infants as young as 8 months are able to remember two label-object pairs after just 5 exposures to each pair. On average, the first looks were significantly oriented more towards the correct puppets, indicating that they familiarize with the labels in a very short time. Longest uninterrupted looks during the test trials are also oriented towards the correct puppets. The majority of infants (n=12) looked to the correct object in both test trials at least in one block, which explains the overall significance of the longest uninterrupted fixations to the correct objects. We therefore conclude that 8-month-old infants are able to learn multiple label-object pairs during a short exposure, as in the present experiment. That is, they expect the correct object to move when named, although the task of switching attention from one label-object pair to another appears to be demanding, so that the infants do not succeed in it in most of the trials.

These results are in partial contrast with those in Experiments 5 and 6, which show that infants do not switch their attention from one auditory regularity to another as easily. Although a further control study would be needed to assess these differences in more detail, we can confirm the initial hypothesis that the cognitive control needed to switch from one auditory regularity to another is different from the cognitive control needed to switch from one word-referent pair to another.

3.4

General discussion

The aim of this chapter was to assess the differences between learning auditory regularities and learning words. More specifically, we were interested in the effect that cross-modality has on learning. Learning new words that map onto the objects or events that surround the language learners is normally an intermodal event that connects the auditory and the visual input into new meaningful units. Although we could argue that acquiring the meaning of words actually implies mapping an abstract auditory form onto an abstract concept (Carey, 2010; Swingley, 2010), the mapping is overwhelmingly initiated and instantiated by concrete visual and auditory exemplars of each category. We can therefore refer to the process of word learning as to cross-modal learning. Recently, infants have also been shown to form cross-modal associations of auditory regularities and visual stimuli (Hochmann et al., 2011; Kovács, 2008). Whereas learning multiple auditory regularities has been shown in infants (Gerken et al., 2011), an interleaved learning of multiple cross-modal associations between auditory regularities and visual stimuli has been proven difficult for infants (Kovács & Mehler, 2009b). We hypothesized that if learning multiple cross-modal associations of auditory regularities and visual stimuli is not demanding because of the possible interferences between two regularities, switching between one association and another might be too demanding for infants. We therefore conducted two experiments in which infants learned to associate two auditory regularities to two positions where the reinforcers appeared, and one word-learning experiment.

In all three experiments, a relatively novel procedure was used that includes gaze-contingent presentation of the stimuli. The reinforcers were appearing conditionally, i.e. when the infants actively looked to the correct part of the screen or the correct object. We decided to use this procedure in order to possibly enhance the learning process. We have observed relatively higher proportion of infants who actively participated compared to other studies (Johnson, Amso, et al., 2003; Kovács, 2008). It also gave us the opportunity to observe the changes in speed of gaze responses. Interestingly, the average time infants needed to look to the correct side of the screen or the correct object was relatively long compared to a more elementary task in which infants learned to associate the static cue to the appearance of the reinforcer (Wang et al., 2012). However, when compared to the average response latencies in similar experiments which used a constant time interval between the auditory stimulus and the reinforcer, the reaction times were not very different (Kovács, 2008). This could indicate that the time needed for infants to make the decision about where to look based on the auditory input is independent of the type of reinforcing and constant, somewhere between 900 and 1000ms.

The results from Experiment 5 and 6, taken together, indicate that infants cannot learn to associate multiple auditory regularities to multiple visual reinforcers. When the two auditory regularities were both from the linguistic domain, the majority of infants (11 out of 20) learned to associate only one of them. Because the stress-based categorization was possible only after a learning phase, one possible explanation is that the stimuli from one category interfered in the other category. We therefore added the control group in which the auditory regularities were distinguishable also at the domain level, one coming from the linguistic and the other from a non-linguistic domain. The control group exhibited the same learning pattern as the previous. Infants thus associate auditory regularities to visual reinforcers equally well both if the two auditory regularities are perceptually easily distinguishable because they come from two different domains (Experiment 6), and if they have to be extracted from the input that belongs to one single domain (extracting the position of lexical stress from speech in Experiment 5).

In contrast to these results, the results of Experiment 7 show that infants are able to learn two label-object pairs in a short period of time and to remember them correctly during the experiment (long-term retention was not assessed). We conclude that learning words requires a different level of cognitive control than associating auditory regularities to visual events. We propose that the difference between learning words and learning regularities is caused by the fact that two different learning processes are involved in these tasks: in the former, infants build new concepts, whereas in the latter, they arbitrarily associate the auditory and the visual input. Labeling enables infants to switch back and forth between two newly built concepts, whereas switching between two arbitrary associations, even if both can be learned and represented, is a more demanding task. This indicates that conceptual knowledge can not only increase infants' memory capacity (Feigenson & Halberda, 2008) but also cognitive control needed to switch between targets. The difference in the responses also confirms the hypothesis that word learning is not a simple associative task as recently proposed (Smith & Yu, 2008; Trueswell et al., 2013; Yu & Smith, 2011).

The studies in Chapters 2 and 3 address the question about the mechanisms of mapping when words are unambiguous, even if there are more than one possible referents for them – which constitutes ambiguity. Natural environment, however, is often ambiguous also as to what constitutes a word because the speech that infants hear is not always clearly delimited into discrete units, namely words. The next chapter will therefore move to this other ambiguity and try to evaluate current hypotheses about what are the possible mechanisms that infants use to extract word candidates from the speech stream they hear.

4

Extracting and remembering plausible word candidates

4.1 Introduction

The problem of word segmentation has been typically built around the question of how (adult) speakers extract possible word candidates from fluent speech and map them onto meaning. Not only for adults, but also for older infants, who already possess some knowledge about their language of exposure, this process could be facilitated by language-specific phonological cues (Cutler & Mehler, 1993; Johnson & Jusczyk, 2001; Langus, Marchetto, Bion, & Nespor, 2012; Millotte et al., 2010). Recently, a relatively large body of research has suggested that tracking statistical regularities could be one of the mechanisms for word segmentation when no language specific knowledge is present, e.g. in very young infants (Aslin et al., 1998; Saffran et al., 1996; Saffran, 2003). Speech, in fact, contains co-occurrence regularities that could indicate word boundaries: across languages, the probability of a syllable to be followed by another syllable (transitional probability, TP) tends to be higher within words and lower at word boundaries (Harris, 1955; Hayes & Clark, 1970). While other cues are usually considered at least as important as statistical ones (Johnson & Jusczyk, 2001; Shukla et al., 2011), in certain cases, transitional probabilities can serve as a sufficient cue for segmenting possible word candidates from continuous speech (Aslin et al., 1998; Thiessen & Saffran, 2003). Although the input in the experimental conditions differs substantially from the input infants normally hear when they learn their first language, tracking transitional probabilities - or possibly some other form of statistical information - could in theory constitute an informative cue also for speech segmentation with natural language input (Ngon et al., 2013; Pelucchi et al., 2009).

Experimental evidence has by and large proceeded without testing the ecological validity of statistical learning in language acquisition: the vast majority of studies with infants use only invariant syllable pairs within words, with each syllable belonging to one word only, so that each syllable can be followed by only one other syllable (TP=1.0), and even in the rare cases where high TPs are compared to the lower word-internal TPs, the results show the preference for the higher TPs (Hay, Pelucchi, Graf Estes, & Saffran, 2011). Such input is very different from the natural language input in which each syllable can be used in multiple words[1], yielding much lower average TP values. We therefore lack the experimental evidence to fully understand how infants process transitional probabilities that are more similar to the probabilities found in natural languages, namely considerably lower than 1.0.

However, let us set this problem aside for the time being, and hypothesize that infants nonetheless pay attention to conditional probabilities in speech. A number of modeling studies aimed to understand the nature of (statistical) information that is present in speech (Gambell & Yang, 2006; Gervain & Guevara Erra, 2012; Swingley, 2005), and to explain the mechanisms that infants may use to segment words from fluent natural-language speech (Batchelder, 2002; Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Swingley, 2005). These models rely on the null hypothesis that tracking statistical co-occurrence regularities in natural language input can lead to preliminary preferences for words. Successful segmentation using statistical cues has been shown in some languages (Gervain & Guevara Erra, 2012; Swingley, 2005), but questioned in others (Batchelder, 2002; Gambell & Yang, 2006). For example, Batchelder (Batchelder, 2002) found substantially better results for English than for either Japanese or Spanish using the same segmentation algorithm.

This contrasts the idea that there are milestones in language development that appear to be universal across languages (Gleitman & Newport, 1995). Despite considerable individual differences among speakers of each language (Weisleder & Fernald, 2013), the approximate timings seem to overlap across languages. One of these is the approximate time in which infants start to produce their first words, around 12 months (Bornstein et al., 2004; Casefli et al., 1995). If statistical co-occurrence regularities constitute a relevant and language-independent piece of information for infants, they should be comparably informative in all natural languages. If languages instead vary in how informative co-occurrence statistics actually are for segmenting possible word candidates, there are two

---

[1] Somewhere between 300 and 2000 different syllables, depending on the language, will constitute a vocabulary of 20000 words for an educated adult. The syllabic repertoire estimates come from our data analyses (see Ch. 4.2), whereas the estimated vocabulary sizes for educated adults vary from 8000 to 40000, 20000 being among the most frequent estimates (Anderson, 1993).

possibilities: either infants rely on other independent cues that indicate whether co-occurrence statistics is informative about word boundaries in their language of exposure, or they do not use probabilistic information to extract word candidates.

There is however another question that arises from the prediction that infants use statistical information to segment words. Performing statistical computations over fragments of the environmental input is thought to enable young infants to acquire and interpret knowledge about the world that surrounds them (Aslin & Newport, 2012; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). This mechanism is usually described as a domain-general associative learning mechanism (Gómez & Gerken, 2000). However, associations do not only help remembering but can also create false memories. For example, participants asked to remember a list of words containing *mad*, *fear*, *hate* and *rage* will not only correctly recall the words *mad* or *fear*. They may also incorrectly admit hearing *anger* that never actually occurred, but is strongly associated in meaning to the memorized words (Roediger & McDermott, 1995). This process, known as associative priming, has important implications that go far beyond incorrectly remembering words on a list (Miller & Gazzaniga, 1998; Yonelinas, 2002). If the brain forms associations between co-occurring environmental stimuli by calculating transitional probabilities (TPs), should statistical calculations also not impair the memory for words? For example, in words like *bikini* and *eskimo* the syllable *ki* can be followed both by *ni* and by *mo*. Because the human mind is much better at calculating adjacent than non-adjacent TPs (Gomez & Maye, 2005; Peña, Bonatti, Nespor, & Mehler, 2002), the first syllable in *bikini* predicts the second syllable *ki* much more strongly than it does the last syllables *ni*. Language learners who rely primarily on adjacent TPs and reconstruct words from possible syllable pairs should thus also remember *phantom words* – such as *bikimo* and *eskini* – that is, statistically probable syllable sequences that never actually occur (see Figure 21). It is even highly probable that language learners could reconstruct great number of such phantom words from the input speech stream, which could possibly create difficulties because it would significantly lower the probability to create a real word. Experiments show that adult participants listening to a continuous sequence of nonsense words do in fact recall *phantom words*, especially if they are more probable than syllable sequences that actually occurred in the stream (Endress & Mehler, 2009). This brings us back to the previous observation that infants were rarely tested for the memory for words with low TPs. It is therefore unknown whether infants would show the same type of memory for phantom words as the adults.

To answer the three questions that follow from the observations above – (1) whether statistical regularities present in child-directed speech are informative for word segmentation, (2) what would be the consequences for the memory load if a language-learner extracted all possible word candidates based on their statistical properties, and (3) whether infants actually remember phantom words as well as true words – three studies were conducted, constituting three parts of the present chapter.

In the first study, we re-evaluate the hypothesis that statistical co-occurrence regularities in all natural languages are informative by analyzing 9 different languages (English, Polish, Dutch, Italian, Spanish, Hungarian, Estonian, Japanese, Tamil) using the same segmentation algorithms that were used in previous studies, transitional probabilities and mutual information (Gambell & Yang, 2006; Swingley, 2005). We also examine in which ways different properties of languages affect their statistical characteristics. In the second study, we estimate the costs and benefits of associative memory in the same child-directed speech corpora that we used the first study. In each corpus, we determined all the possible phantom words that would survive statistical segmentation (Gambell & Yang, 2006; Gervain & Guevara Erra, 2012; Swingley, 2005). We reasoned that if learners rely on statistical regularities alone, they must remember these highly probable phantoms better than the average word. In the third study, we tested 6-month-old infants with the Head-Turn-Preference Procedure (HTPP). We tested infants' memory for the words they heard during the experiment, and compared it to their memory for phantom words.

4.2 Study 1

Co-occurrence statistics as a language dependent cue for speech segmentation

In the following study, co-occurrence probabilities were calculated in child-directed speech corpora from 9 languages. We measured how successful the word segmentation algorithm was in each language and across the 9 languages. Since the speech infants hear is different from adult-directed speech (Thiessen et al., 2005), we used child-directed speech corpora as the input for the segmentation algorithms. While it is clear that a real language learner uses other cues as well, and has various memory and processing limitations, we restricted our word segmentation task to the use of statistical cues. The decision was motivated by the belief that the integration of different cues is only justified once we know how each of the cues contributes to the process individually.

4.2.1 Methods

4.2.1.1 Corpora

We analyzed the transcribed spoken corpora in 9 different languages from the CHILDES database (MacWhinney, 2000): Estonian, Hungarian, Japanese, Tamil, Italian, Spanish, Dutch Polish, and English. Parts of the following corpora were used in each language: Estonian (Argus, 2004; Kohler, 2004; Korgesaar, 2011; Vija, 2004), Hungarian (Gervain & Guevara Erra, 2012), Japanese (Ishii, 1999; Oshima-Takane, Y. MacWhinney, Sirai, Miyata, & Naka, 1998; Ota, 2003), Tamil (Narasimhan, 2004), Italian (Antelmi, 2004; Antinucci & Parisi, 1073; Tonelli, 2004; Volterra, 1976), Spanish (Goga, 2006; Jackson-Maldonado & Thal, 1994; Vila, 1990), Dutch (Bol, 1995; Van Kampen, 1994; Wijnen, 1992), Polish (Smoczynska, 1985; Weist & Witkowska-Stadnik, 1986), and English (Korman, 1984; Swingley, 2005). We chose languages that belong to different linguistic families (Slavic, Romance, Germanic, Finno-Ugric, Dravidian, Japonic) and differed in a number of grammatical features, such as word order, morpho-syntactic complexity and phonological features (Dryer & Haspelmath, 2011). The choice of the languages was finally determined by the availability of the child directed corpora and the availability of native speakers who could segment the corpora into syllabic sequences. In each corpus, only the child-directed sentences spoken by adults were taken into account. Phonetic transcription was used wherever the spelling differed from the orthographic transcription. The syllabified corpora we used for the analysis are available upon request to the authors.

In order to assure that we were analyzing a comparable amount of data, we selected 3300 sentences for each language. Although the size of the corpora typically does not significantly change the co-occurrence statistics (Gambell & Yang, 2006), the relatively small sizes of the corpora could affect the results of the segmentation process. The results in the corpora in our study were therefore compared to the results using the larger amount of input data in Hungarian and Italian available in CHILDES database (Gervain & Guevara Erra, 2012; http://childes.psy.cmu.edu/derived/), and the syllabified corpora of Dutch (Swingley, 2005; van de Weijer, 1998) and English (Korman, 1984; Swingley, 2005), obtained by request to the author. The sizes of the corpora are 15200, 10470, 10700, and 12800 sentences, respectively.

The languages we chose differ significantly in a number of quantitative features, such as average word and utterance length and syllabic diversity (Table 4). In all languages, co-occurrence statistics are stronger within words and weaker at word boundaries: average values in all dependencies measures are lower in the across-word syllable pairs and higher in the within-word syllable pairs (Table 5). This confirms the findings of previous studies where TP drops at word boundaries were observed.

	English	Polish	Dutch	Spanish	Italian	Hungarian	Estonian	Japanese	Tamil
Words	9196	16529	14475	14710	15777	12669	14590	9621	9226
Mean utt. lenght (words)	3.97	6.38	5.37	5.48	5.79	4.69	4.83	3.99	3.80
Mean utt. lenght (syll)	4.55	8.71	5.60	7.44	8.75	6.51	7.47	6.98	6.55
Mean word length (syll)	1.27	1.74	1.28	1.67	1.83	1.70	1.69	2.33	2.34
% words longer than 2 syll	0.01	0.16	0.05	0.11	0.21	0.15	0.13	0.35	0.36
Syllables types	2881	1137	1049	764	788	1718	1289	371	1108
Bigrams types	7720	8936	8200	7474	7852	10807	10710	3956	7108
Bigrams tokens	11703	28738	18467	24571	28882	21492	24649	23029	21612

Table 4 Quantitative features of the corpora. The abbreviation utt. stands for utterance and syll. for syllable.

		mean word-internally	mean word-straddling	difference
English	FTP	0.493	0.156	0.337
	BTP	0.477	0.158	0.320
	MI	-5.295	-10.196	4.901
Polish	FTP	0.145	0.126	0.019
	BTP	0.259	0.080	0.180
	MI	-10.691	-12.554	1.863
Dutch	FTP	0.387	0.114	0.273
	BTP	0.321	0.126	0.195
	MI	-7.057	-11.271	4.215
Spanish	FTP	0.181	0.092	0.089
	BTP	0.233	0.072	0.161
	MI	-10.194	-12.674	2.480
Italian	FTP	0.242	0.041	0.202
	BTP	0.133	0.093	0.040
	MI	-10.979	-13.196	2.218
Hungarian	FTP	0.228	0.153	0.075
	BTP	0.323	0.108	0.215
	MI	-7.454	-11.057	3.603
Estonian	FTP	0.233	0.096	0.137
	BTP	0.186	0.112	0.074
	MI	-9.520	-11.578	2.058
Japanese	FTP	0.141	0.056	0.085
	BTP	0.101	0.088	0.013
	MI	-12.881	-14.063	1.182
Tamil	FTP	0.262	0.109	0.153
	BTP	0.222	0.141	0.081
	MI	-8.954	-11.717	2.763

Table 5 Average values in all dependency measures (forward and backward transitional probabilities, and mutual information) for word-internal and word-straddling syllable pairs.

4.2.1.2 Dependency measures

We analyzed adjacent dependencies (forward transitional probabilities (FTP), backward transitional probabilities (BTP), and mutual information (MI) among the syllables in our corpora. The dependencies were computed as follows:

FTP(XY) = frequency(XY) / frequency(X)

BTP(XY) = frequency(XY) / frequency(Y)

MI(XY) = $\log_2$ (frequency(XY)/(frequency(X) * frequency(Y)))

Adjacent transitional probability (either forward or backward) is the conditional probability statistics that measures how predictive adjacent elements are. It is the main statistical measure in various word segmentation models (Aslin et al., 1998; Frank et al., 2010; Tyler & Cutler, 2009). Mutual information is a symmetrical measure, similar to transitional probabilities. It has been used to measure the strength of associations between words in written corpora and is now used in many corpora for extracting frequently co-occurring word pairs (Hayes & Clark, 1970). Recently, mutual information has also been used to model the word segmentation process (Swingley, 2005). This measure is usually not normalized, thus its range varies in different corpora.

Non-adjacent dependencies were not measured because in our corpora the proportion of words containing more than two syllables is relatively low (Table 4). Furthermore, infants appear to disregard the non-adjacent dependencies when the adjacent ones are high (Gomez, 2002; Gomez & Maye, 2005). Although smaller and larger constituents are sometimes considered as well (Batchelder, 2002; Bonatti, Peña, Nespor, & Mehler, 2007; Brent & Cartwright, 1996; Goldwater, Griffiths, & Johnson, 2007), the syllable has been predominantly recognized as the minimal perceptual unit in speech and used as the minimal input unit also in models of infant speech segmentation (Bertoncini, Floccia, Nazzi, & Mehler, 1995; Gambell & Yang, 2006; Mehler, 1981; Saffran et al., 1996; Swingley, 2005). We therefore used adjacent dependencies among syllables in our study.

4.2.1.3 Segmentation algorithms

In each of the corpora we used two segmentation algorithms to determine possible word boundaries, both using only the dependency measures described above. In the first algorithm (Relative), the boundaries are set wherever the dependency measure (FTP, BTP, or MI) of a syllable pair (XY) is weaker than in the neighboring ones (ZX and YW), defined as:

$$FTP(ZX) > FTP(XY) < FTP(YW).$$

This segmentation algorithm has been proposed as the mechanism used by infants in order to segment words from artificial speech streams and is usually seen as a psychologically plausible mechanism (Aslin et al., 1998; Saffran et al., 1996).

But since such an algorithm can not (by definition) find any mono-syllabic words (Gambell & Yang, 2006), we used also an algorithm that looks for general drops of TP values below a certain threshold and sets the word boundaries there (Absolute). Although in principle many different thresholds can be used for delimiting words in the stream (Gervain & Guevara Erra, 2012; Swingley, 2005), it is highly improbable that any real language learner would repeatedly segment word candidates from the input using many different thresholds and then select the threshold that gives the best result. Furthermore, when we compared the threshold that gave best results among 100 percentile thresholds (Gervain & Guevara Erra, 2012) to the results using the average values of the word-internal syllable pairs, we found no significant difference ($t(16)=0.8615$, $p=0.404$). We therefore took the latter as a suitable absolute threshold.

4.2.1.4 Evaluation measures

We evaluate each segmentation strategy using the conventional information retrieval measures (Baeza-Yates & Ribeiro-Neto, 1999): Precision, Recall, and their harmonic mean, F-score, as defined in:

$$\text{Precision} = \#\text{hits}/(\#\text{hits}+\#\text{false alarms})$$
$$\text{Recall} = \#\text{hits}/(\#\text{hits}+\#\text{misses})$$
$$F_1 = (2*\text{precision}*\text{recall})/(\text{precision}+\text{recall})$$

Whenever the proportion of hits is substantially lower than the proportion of falsely selected or missed words (when either precision or recall are lower than 40%), the F-score will be lower than 0.5. It is therefore generally accepted that the F-scores lower than

0.5 indicate relatively noisy word extraction (Makhoul, Kubala, Schwartz, & Weischedel, 1999).

4.2.1.5 Procedure

In each corpus, we first measured the dependencies (FTP, BTP, MI) among syllables within words and at word boundaries. In a second step, the information about the word boundaries was removed and only the utterance boundaries remained. Information about utterance boundaries remained because utterances in child directed speech corpora are – according to the transcription rules – delimited either by pauses or by the utterances of someone else. In both cases a word boundary is indicated without any doubt. Each of the segmentation algorithms produced a distinct set of word candidates in each language. These word candidates were compared to the actual words in the same corpus. The input for learning (measuring the dependencies) and modeling the word segmentation was the same because we wanted to directly compare the actual words and the word candidates in each corpus.

The scripts used to analyze the corpora, to extract the word candidates from the stream with the two algorithms, and to evaluate the results, are available in the Appendix 1.

4.2.2 Results

The comparison between the three dependency measures showed that in the Dutch, Italian, Estonian, Japanese, and Tamil corpora, forward TPs give better results than backward TPs, while in Hungarian and Polish, backward TPs are more informative in both segmentation algorithms (Table 6; Figure 19; Table 7; Figure 20). It was recently suggested that possible reasons for the differences could lay in the morpho-syntactic differences between the analyzed languages: in head-initial languages with a default subject-verb-object (SVO) word order, such as English or Italian, FTPs could be more informative than BTPs. In head-final languages with a default SOV word order, BTPs tend to be more informative (Gervain & Guevara Erra, 2012; Onnis & Thiessen, 2013). Our language samples cannot confirm this hypothesis. BTPs are more informative only in Polish, which is a head-initial language with a default SVO order, and in Hungarian, whereas in other head-final languages in our sample, Japanese and Tamil, FTPs and MI are more informative. We find no correlation between the directionality of phrase structure/word order and the co-occurrence statistics across languages (linear correlation coefficient (corrval)=-0.05, p=0.786). When the differences between the three measures are compared across languages, they turn out not to be significant, at least in the analyzed corpora (two-way ANOVA with the factors algorithm and measure showed no differences between the results in each of the measures: $F(2,50)=0.12$, $p=0.89$). The effect of the algorithm was instead significant: $F(1,50)=8.28$, $p=0.006$, with no interaction between the factors: $F(2,50)=0.04$, $p=0.9633$). For further analysis we therefore collapsed the results in the three measures.

In order to verify the possibility that the results we obtained were influenced by the relatively small sizes of the corpora, we compared the results in the smaller corpora to the results obtained from larger corpora, where the larger corpora were available. In Hungarian, Italian, and English, our corpora were the subsets of the larger corpora, whereas in Dutch, two distinct corpora were compared. Paired two-tailed t-test between the larger and shorter input showed no significant differences in Italian, Hungarian, and Dutch ($t(5)<2.65$; $p>0.05$), whereas in English there was a difference when using the Relative algorithm ($t(2)=8.79$, $p=0.01$), but not in the Absolute algorithm.

In Italian, Estonian, Hungarian, Dutch and English, the Absolute segmentation algorithm produced relatively accurate word lists, with the F-scores higher than 0.5, and significantly higher than in the Relative algorithm ($t(2)>4.61$, $p<0.044$) (Table 6). The Dutch and English corpora were segmented with very high accuracy in the Absolute algorithm ($F_1>0.77$ in all measures). This could be due to the high proportion of monosyllabic words: if on average, word-straddling dependencies are relatively low and there are only a few longer words with relatively higher internal dependency values, the only real falsely selected word candidates are frequently co-occurring monosyllables. In fact, when we tried to segment word candidates by putting a word boundary after every syllable, Dutch and English were the only languages among the ones we analyzed in which such segmentation was successful (for English, $F_1=0.79$, for Dutch, $F_1=0.75$). In Tamil and Japanese, word segmentation with the Relative algorithm reaches the F-scores higher than 0.5 and is significantly better than with the Absolute algorithm (Tamil: paired two-tailed t-test, $t(2)=4.98$, $p=0.038$; Japanese: $t(2)=11.14$, $p=0.008$) (Table 7). Word extraction is relatively weak in both segmentation algorithms in Spanish and in Polish; with the highest F_1-scores lower than 0.5 (the proportion of correctly selected words did not exceed the proportion of falsely selected and missed words). In Spanish, the best result is F_1=0.41 in the Absolute algorithm when FTP are measured, and in Polish, F_1=0.42 in the Absolute algorithm when BTP are measured. Judging from our results, the statistical segmentation of words is noisy in both Spanish and Polish.

The presented overall F-scores are harmonic means of the proportions of correctly extracted words compared both to the words that were missed and to the falsely selected chunks of speech. The nature of falsely selected chunks of speech could be further informative about the difficulties that language learners can face in remembering these chunks. That is, if the miss-segmented chunks are composed of two or more full words, infants would be over-segmenting the stream, but the edges would be set correctly. For example, connected speech in Italian is segmented with Relative algorithm as follows:

La-cit-ta.
Tu-non-mi pre-pa ri-mai-nien-te.
Se-bat-ti-la tes-ta sul-pa vi-men-to ti-fai-ma le.
Do-po la-mam-ma to-glie-il-co per-chio-di nuo-vo.
Qua-li so-no-le me-di-ci ne-che mi-da la-mam-ma.

Correct segmentation would be:

La cit-ta
Tu non mi pre-pa-ri mai nien-te.
Se bat-ti la tes-ta sul pa-vi-men-to ti fai ma-le.
Do-po la mam-ma to-glie il co-per-chio di nuo-vo.
Qua-li so-no le me-di-ci-ne che mi da la mam-ma.

Some words are chopped in the middle by the Relative algorithm, such as in the sequence *sul-pa vi-men-to ti-fai-ma le*. But there are also some segmented parts that leave the word boundaries intact: *se-bat-ti-la tes-ta*, where the first chunk is made of three words. Although these are both counted as false segmentations, the second type of mistakes can be possibly corrected with more exposure. In addition, in data on infant speech production, the latter type of miss-segmentation is attested, but not the former: toddlers and children, albeit rarely, both under- and over-segment the speech stream, but never glue together pieces of adjacent words (Cutler, 2012; Pinker, 1994). The number of over-segmentations is bigger than the number of under-segmentations because of phrases, idioms, and frequently collocated words (such as *IwannaX* in English). We therefore computed the proportions of over-segmentations compared to other miss-segmented chunks in the lists of word candidates segmented by each of the two algorithms. The results show that in the Relative algorithm, which cannot segment chunks smaller than two syllables, the proportions of over-segmented parts are relatively high, spanning from 15.1% in Tamil to 85.8% in English, both with BTPs measured. In the Absolute algorithm, the proportions are much lower, from 41.3% in English to 2.7% in Tamil, both with FTPs measured (see Table 8). In the Absolute algorithm, the languages with better overall results also have a higher proportion of over-segmented words (rho=0.53, p=0.007), whereas there is no such correlation in the Relative algorithm (rho=-0.36, p=0.08).

A closer inspection revealed that languages with better results in each of the algorithms share an important linguistic feature: the two languages with the best results in the Absolute algorithm belong to the so-called stress-timed rhythmic class, and the two languages that have the best results in the Relative algorithm belong to the so-called mora-timed rhythmic class (Nespor, Shukla, & Mehler, 2011; Ramus, Nespor, & Mehler, 1999). It is possible that rhythmic properties are associated with the statistical properties of the languages we analyzed. We therefore compared the results in the analyzed languages to the average proportions of vocalic intervals – the measure proposed in Ramus et al 1999 – in the same languages. The average proportions of vocalic intervals were measured on independent auditory corpora and were shown to correlate significantly to the perceived linguistic rhythm (Nespor et al., 2011; Ramus et al., 1999). The measurements exist for all the analyzed languages except Estonian, which was consequently excluded from the comparison. We found that both algorithms correlate to the rhythm measurements: word segmentation results with the Relative algorithm are positively correlated to the proportion of vocalic intervals (corrval=0.636, p=0.0008; Figure 20), whereas the segmentation with the Absolute algorithm is negatively correlated to the proportion of vocalic intervals (corrval=-0.458, p=0.025; Figure 19). This suggests that at least some of the variation in statistical segmentation can be explained by rhythmic differences between the languages we analyzed.

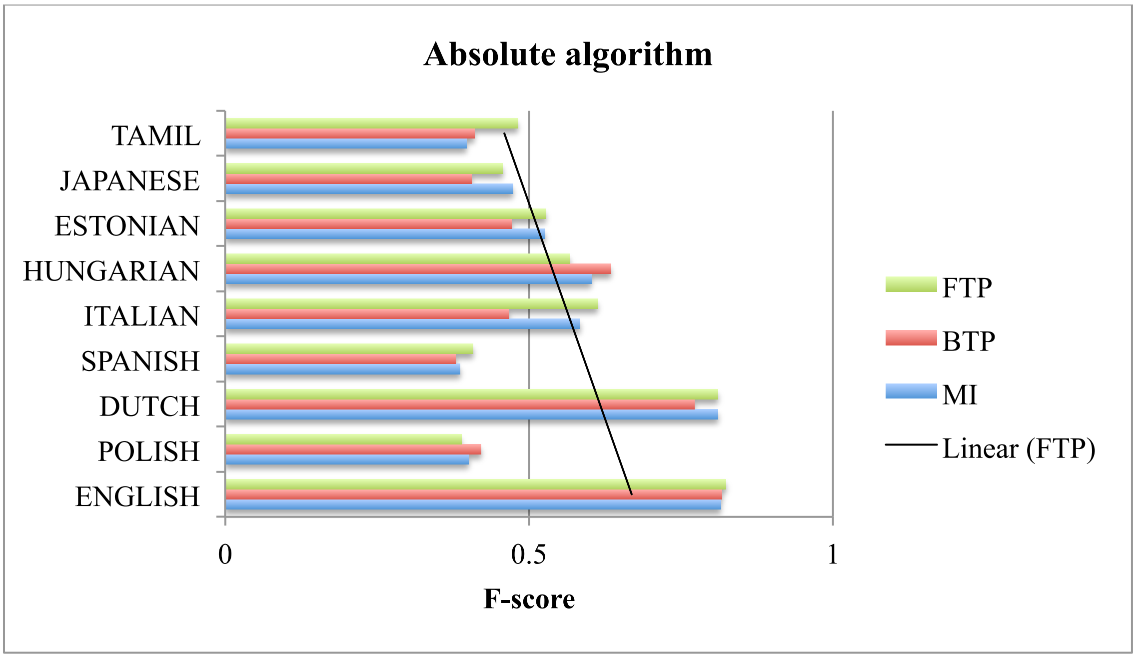

Figure 19 The results of the segmentation using the Absolute algorithm. The languages are listed according to their average percentage of vocalic intervals (the measure for linguistic rhythm). The line represents the linear regression function.

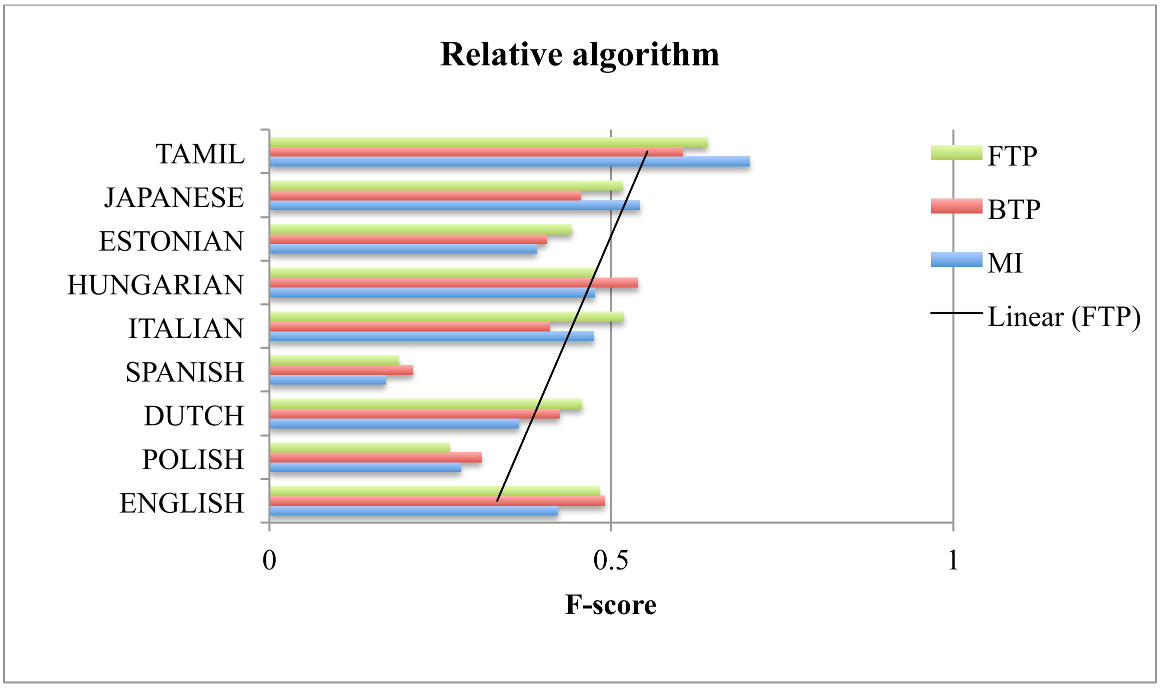

Figure 20 The results of the segmentation using the Relative algorithm. The languages are listed according to their average percentage of vocalic intervals (the measure for linguistic rhythm). The line represents the linear regression function.

	Absolute algorithm								
	FTP			BTP			MI		
	recall	**precision**	**F-score**	**recall**	**precision**	**F-score**	**recall**	**precision**	**F-score**
English	0.82	0.83	0.82	0.81	0.83	0.82	0.85	0.78	0.82
Polish	0.45	0.34	0.39	0.50	0.37	0.42	0.45	0.36	0.40
Dutch	0.83	0.79	0.81	0.79	0.76	0.77	0.85	0.78	0.81
Spanish	0.45	0.37	0.41	0.42	0.34	0.38	0.41	0.36	0.39
Italian	0.72	0.54	0.61	0.54	0.41	0.47	0.63	0.54	0.58
Hungarian	0.63	0.51	0.57	0.70	0.58	0.63	0.67	0.54	0.60
Estonian	0.59	0.48	0.53	0.52	0.43	0.47	0.57	0.49	0.53
Japanese	0.59	0.37	0.46	0.51	0.34	0.41	0.54	0.42	0.47
Tamil	0.59	0.41	0.48	0.51	0.34	0.41	0.49	0.33	0.40

Table 6 Table of the results in Absolute algorithm in all languages with the three measured dependency measures. Recall, precision, and their harmonic mean (F-score) are measured as defined above (4.2.1.4).

	Relative algorithm								
	FTP			BTP			MI		
	recall	**precision**	**F-score**	**recall**	**precision**	**F-score**	**recall**	**precision**	**F-score**
English	0.40	0.61	0.48	0.41	0.62	0.49	0.33	0.58	0.42
Polish	0.24	0.30	0.26	0.28	0.35	0.31	0.24	0.34	0.28
Dutch	0.38	0.57	0.46	0.35	0.54	0.42	0.28	0.51	0.36
Spanish	0.17	0.22	0.19	0.18	0.24	0.21	0.14	0.21	0.17
Italian	0.47	0.58	0.52	0.37	0.46	0.41	0.41	0.57	0.47
Hungarian	0.43	0.54	0.48	0.48	0.62	0.54	0.40	0.59	0.48
Estonian	0.40	0.50	0.44	0.36	0.46	0.41	0.33	0.48	0.39
Japanese	0.53	0.50	0.52	0.47	0.44	0.45	0.53	0.56	0.54
Tamil	0.64	0.64	0.64	0.60	0.61	0.60	0.67	0.74	0.70

Table 7 Table of the results in Relative algorithm in all languages with the three measured dependency measures. Recall, precision, and their harmonic mean (F-score) are measured as defined above (4.2.1.4).

		Proportion of super-segmentations among incorrectly segmented words								
		English	**Polish**	**Dutch**	**Spanish**	**Italian**	**Hungarian**	**Estonian**	**Japanese**	**Tamil**
Absolute algorithm	**FTP**	0.41	0.11	0.23	0.12	0.04	0.09	0.11	0.04	0.03
	BTP	0.36	0.08	0.26	0.09	0.08	0.10	0.11	0.06	0.03
	MI	0.08	0.19	0.10	0.19	0.25	0.09	0.14	0.22	0.03
Relative algorithm	**FTP**	0.85	0.37	0.81	0.54	0.54	0.47	0.57	0.21	0.18
	BTP	0.86	0.57	0.79	0.64	0.37	0.62	0.46	0.17	0.15
	MI	0.80	0.68	0.84	0.80	0.68	0.79	0.67	0.32	0.43

Table 8 The proportion of over-segmented chunks compared to other miss-segmented word candidates in both algorithms.

4.2.3 Discussion

The aim of this first study was to examine how informative co-occurrence statistics is for segmenting words from the un-segmented input in various languages when no other cue is present. We analyzed child-directed corpora from 9 languages that differ not only at the morpho-syntactic level, but also at the phonological level, the aspect of language that young infants are highly attentive to. Co-occurrence statistics can be measured with different measures; in this study we use (forward and backward) transitional probabilities and mutual information. The word boundary was set wherever the dependency value dropped below the average word-internal value (Absolute algorithm) or wherever the value was lower than the neighboring ones (Relative algorithm). Word lists obtained with these algorithms were then compared to the real words in the same corpora (using an information retrieval measure, F-score).

In Polish and Spanish, segmentation is noisy – the proportion of correctly selected words does not exceed the proportions of falsely selected and missed words. This remains unchanged even if we count all over-segmented chunks as correctly segmented words: the F-scores remain lower than 0.5 (in Polish, F_1=0.44 in Absolute algorithm with FTPs measures; in Spanish, F_1=0.49 in Relative algorithm when BTPs are measured). For Spanish, Batchelder (Batchelder, 2002) proposed that decreased word- and utterance-length could improve the results of statistical word segmentation. But when our Spanish sample was compared to both longer and shorter samples of Italian – which belongs to the same rhythmic class and where the results are considerably better (Italian: F_1=0.61, Spanish: F_1=0.41 (in Absolute algorithm)) – Spanish did not differ significantly from Italian in any of the analyzed features (average word- and utterance-length and the proportion of monosyllabic words), except for the difference between the word-internal and word-straddling FTPs, which is higher in Italian (0.20) than in Spanish (0.03) (see Table 5). It appears that words in Italian are on average statistically more informative than in Spanish, at least in our corpora. For Polish, however, it has been suggested that it could be a representative of a fourth rhythmic class because of its mixed properties, such as low proportion of vocalic intervals and high word-length (Ramus, Dupoux, & Mehler, 2003). In our study, word-internal dependencies in Polish are much lower than in other stress-timed languages (Table 5), and the results in the Absolute algorithm are low. It is possible that statistical dependencies are not very informative for the rhythmic class that

Polish might represent. To further investigate this possibility, other languages belonging to the same class would have to be analyzed.

In other languages in our sample, at least one of the two segmentation strategies we used is relatively informative – the proportion of selected words exceeds the proportion of missed or falsely selected words. In the Absolute algorithm, the languages that have the best overall results (English ad Dutch) also have the highest proportion of over-segmented chunks, possibly indicating even better word recognition in general. Our results for English confirm higher results of segmentation tasks reported in other studies (Batchelder, 2002; Fourtassi, Orschinger, Dupoux, & Johnson, 2013).

The results, moreover, correlate significantly to the rhythmic classes the different languages belong to. The Absolute algorithm was most successful in the languages with the lowest proportion of vocalic intervals, and the Relative algorithm was successful only in the languages with a very high proportion of vocalic intervals. There is at least one possible explanation for this correlation: languages with a lower proportion of vocalic intervals in our sample have shorter words (corrval=0.698, $p<0.0001$; Nespor et al., 2011) and languages with shorter words tend to have higher word internal dependency values (corrval=-0.296, $p=0.049$). The algorithm that finds the words according to the absolute values of the dependencies (Absolute algorithm) is therefore able to extract many monosyllabic words as well as polysyllabic words with high internal dependency values, yielding a relatively successful overall word extraction rate. The algorithm that finds the word boundaries by looking for the local dips (Relative algorithm) is, instead, unable to extract monosyllabic words, but is more sensitive to the differences between word-internal values and values at the edges of words. It is therefore more successful in the languages with long words, such as the so-called mora-timed languages.

Could infants take advantage of this information when segmenting words? They probably cannot estimate the average word length if the process of word segmentation is still ongoing, but they are sensitive to the rhythm of their input language from birth on (Bertoncini et al., 1995; Nazzi et al., 1998; Peña, Pittaluga, & Mehler, 2010; Ramus et al., 2000). The identification of the rhythmic class of their language of exposure could offer a bias as to the most common word length: mean word length and syllabic complexity are, in fact, inversely proportional (Mehler & Nespor, 2004). Our results show that languages with a distinct (stress- or mora-timed) rhythm can be segmented relatively well using only statistical information, while languages with the syllable-timed rhythm less so. It could therefore be possible that infants can modify the way they use statistical regularities (among other cues) to segment words according to the rhythmic class of the language they are exposed to: infants who hear a language with a very distinct rhythm, such as Tamil or English, will pay more attention to co-occurrence regularities than they will in a language with a less coherent rhythmical signature, such as Hungarian, Estonian, or Polish.

Such a prediction makes co-occurrence statistics a language-dependent, non-universal cue for word segmentation, similar to phonological cues such as lexical stress-patterns (Jusczyk, Houston, & Newsome, 1999; Mattys, Jusczyk, Luce, & Morgan, 1999). This would imply that statistical learning should be set as any other linguistic parameter, which is not a very plausible proposal. Namely, one of the proposed advantages of using statistical regularities in language acquisition is that they represent low-level and language independent cues. It was therefore proposed that they can be used before language-specific cues for word segmentation (Swingley, 2005; Thiessen & Saffran, 2003), which is in direct opposition to the prediction presented above. Our study cannot address the question whether statistical cues are actually used by infants during word segmentation and which are the cues infants rely mostly on. Our results, however, show that the amount of statistical information present in various languages varies considerably. Further studies on statistical learning with infants will therefore not only have to show that infants use transitional probabilities for segmenting artificial speech, but also account for how young infants may deal with these cross-linguistic differences that suggest that statistical cues in speech segmentation are language-specific.

4.3 Study 2

On the limits of associative memory

Let us now hypothesize than an infant listens to the language in which co-occurrence statistics seems to be informative about word boundaries. If such an infant really tracks the transitional probabilities among the adjacent syllables, it is possible that many highly probable syllabic sequences – that never occur in the input – are memorized together with the actual words from the input. Before examining whether infants actually remember such phantom-like word candidates in an experimental situation, we analyze again the corpora that we used during the statistical segmentation task. We model the situation in which a hypothetical learner extracts all highly probable polysyllabic sequences – both the actual words and the phantom sequences.

4.3.1 Methods

We used the same corpora as in the previous study and performed the same initial analysis as before (see 4.2.1.1). In order to maintain greater clarity we only focused on the FTPs (and did not perform the analysis also with BTPs and MI), hypothesizing that the results with other dependency measures would be similar. We calculated FTPs for all syllable pairs in the corpora, and divided the syllable pairs into those that occur within words (TP_{within}) and those that straddle word boundaries (TP_{across}). In order to determine and memorize a highly probable syllable pair that could constitute a word, a possible language learner might determine a threshold above which a syllable pair is determined to be a word-internal one (see Ch. 4.2.1.3). There are many possible thresholds, and no unanimous agreement as to which is the strategy used by human learners. We therefore used the same threshold that was used in Study 1, the average TP_{within} values. In this way, we obtained a group of syllabic pairs that are probably word-internal in each language.

We then calculated a) how many probable syllable pairs actually constitute bi-syllabic words, and b) how many of the possible combinations of syllabic pairs into 3- or 4-syllabic sequences constitute real 3- and 4-syllabic words in the input (see 4.1; Figure

21). All the probable bi-, 3-, or 4-syllabic sequences that were not real words were labeled as "highly probable phantom words" (Endress & Mehler, 2009).

The script used to analyze the corpora and extract the phantom words is available in the Appendix 2.

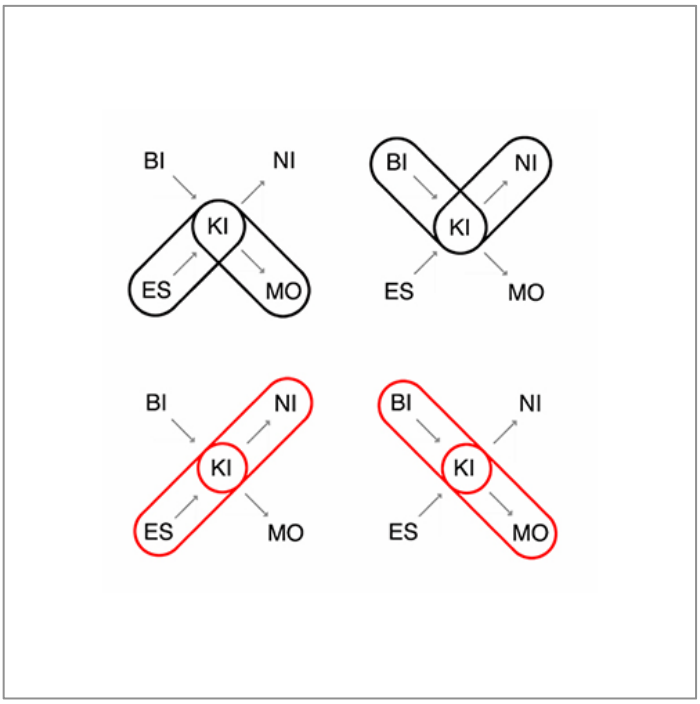

Figure 21 The formation of phantom words. The syllable pairs that form words like *eskimo* and *bikini* – *eski, kimo, biki* and *kini* – could also form phantom words such as *eskini* and *bikimo*.

4.3.2 Results

Figure 22 shows the proportions of words compared to phantom words in each of the nine analyzed languages. The absolute numbers are presented in Table 9. When we compare the words and the phantom words formed of all the syllable pairs that are statistically more probable than the average word, only about half of the bi-syllabic sequences are words (M=0.523, SD=0.072). If infants rely on TPs, they must thus be at chance in discriminating bi-syllabic words from bi-syllabic phantoms. The proportion of words decreases drastically in three-syllabic (M=0.013, SD=0.006) and four-syllabic sequences (M=0.0004, SD=0.0004; see Figure 22). Correctly remembering longer words by calculating statistical regularities between syllables must thus be close to impossible.

To grasp the scope of the problem, we compared the amount of statistically probable phantom words to the vocabulary of an average college educated adult, which has been estimated to be around 25,000 words (D'Anna, Zechmeister, & Hall, 1991). Few hours of child-directed speech will contain on average almost 60 times more highly probable phantom words than the total number of words an individual will learn throughout his lifetime (M=58.97 times more phantom words; SD=55.3). Even if we only consider 2- and 3-syllabic sequences, which correspond to the length of the vast majority of actual polysyllabic words in the nine languages we considered (M=0.858, SD=0.10), the number of highly probable syllabic sequence still exceeds the average lexicon size (M=50290 phantom words, SD=29857).

However, similarly to the results of Study 1, we find substantial differences among the analyzed languages. When we compared the proportion of words among highly probable syllable sequences, we again find similar results as in Study 1 – the proportion of words is correlated to the linguistic rhythm (measured with the proportion of vocalic intervals in speech) that a language belongs to (corrval=0.768, p=0.0261). In Japanese, for example, there are "only" 22 times as many phantom sequences as real words, whereas in Polish, there are almost 55 times as many phantoms as words (if we take only the 2- and 3-syllabic sequences). The problem of selecting words from all the possible sequences should be therefore considerably more difficult in Polish than in Japanese. This is surprising because children from different linguistic environments are known to achieve the major developmental milestones in language acquisition in comparable time (Gleitman & Newport, 1995).

	2-syllabic words	**highly probable bisyllables**	**3-syllabic words**	**highly probable 3-syllables**	**4-syllabic words**	**highly probable 4-syllables**
TAMIL	785	886	869	43676	440	760200
JAPANESE	461	264	576	22788	415	851957
ESTONIAN	1213	864	586	70679	253	1786544
HUNGARIAN	1271	1746	841	66274	254	1462637
ITALIAN	720	426	800	51904	354	1775083
SPANISH	805	798	469	37383	131	1311261
DUTCH	513	564	233	22033	44	40869
POLISH	1267	1494	853	114981	188	4822496
ENGLISH	393	341	59	15536	8	5370

Table 9 The total number of words and highly probable syllable combinations in each of the analyzed corpora.

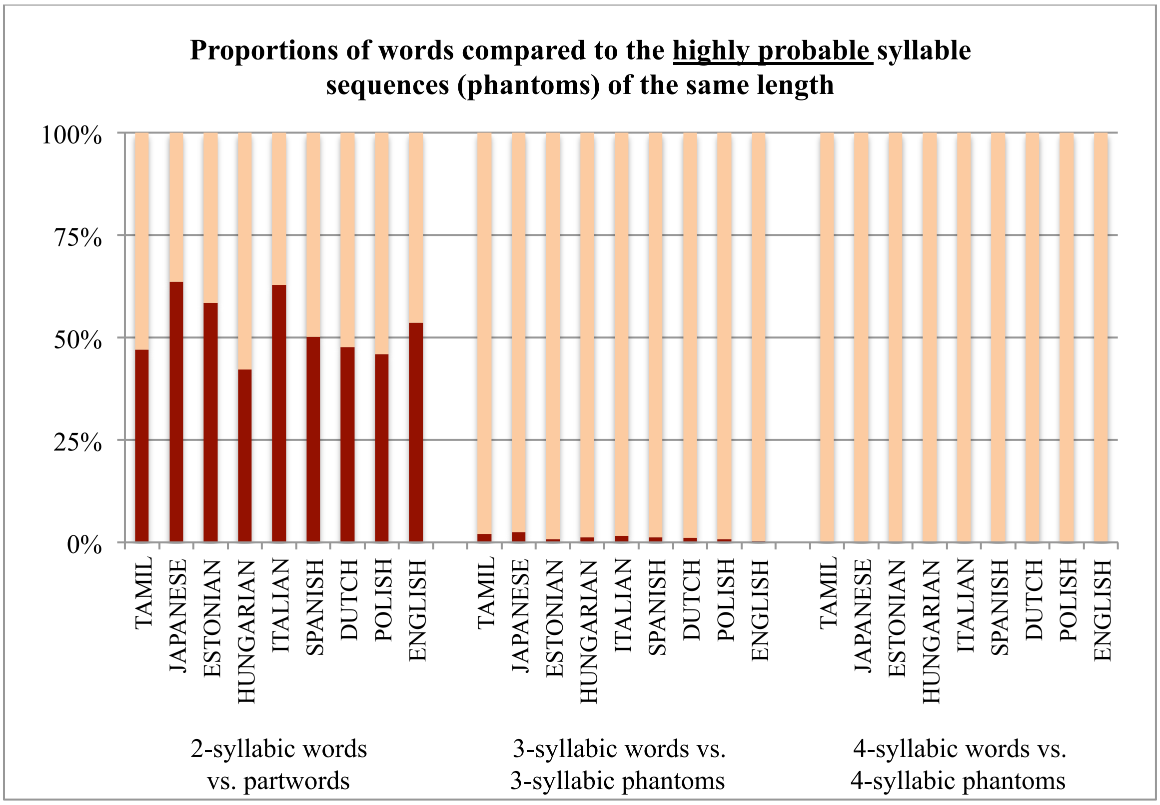

Figure 22 Proportion of words compared to syllable sequences that are more probable than the average syllable sequence within words.

4.3.3 Discussion

Our results show that phantom words – statistically probable but unheard syllable sequences – are more numerous than actual words. The longer the words are, the less predictive the co-occurrence probabilities become. If statistical learning enables young learners to extract relevant information from the environmental input, but also causes false memories of words they never heard, unconstrained statistical learning will cause recollections of phantom words with such a high probability that acquiring a language should in fact be impossible. Unconstrained statistical learning thus violates some of the core facts about language acquisition. Linguistic input is clearly necessary for a language to be acquired because children will only learn the language that they are exposed to. However, if infants remember more nonsense words than actual words, the linguistic input can no longer be informative. Finally, the cross-linguistic variation in the proportion of phantom words would also hinder children in different linguistic environments to achieve the developmental milestones in comparable time. It should cause cross-linguistic differences in language acquisition that have actually never been observed (Gleitman & Newport, 1995).

Clearly, TPs between adjacent syllables are not the only distributional cues present in the speech stream. Besides adjacent TPs, infants calculate some non-adjacent TPs between syllables, possibly as early as by 3-months of age (Mueller, Friederici, & Männel, 2012) and Bayesian modeling suggests that combining distributional information from different sources could improve speech segmentation (Goldwater, Griffiths, & Johnson, 2009). However, young infants appear to calculate non-adjacent dependencies only when the adjacent TPs are low (Gomez & Maye, 2005). As the number of probable but unattested syllable sequences could even increase if both adjacent and non-adjacent TPs are considered, we leave the question about what would the possible consequences of tracking both adjacent and non-adjacent probabilities simultaneously for possible future research. Further, the strength of the association between adjacent syllables will not only depend on the probability but also on the frequency of co-occurrence. For example, two syllables that co-occur only with each other will have a TP of 1.0 regardless whether they occur only once or a thousand times. Yet young infants are known to recognize highly frequent phantom-like syllable sequences that span word boundaries better than actual words that are less frequent (Ngon et al., 2013), showing that the frequency of occurrence is likely to affect memory. This suggests that the phantom word problem persists across different statistical computations and that combining diverse distributional cues is unlikely to narrow the space of possible word candidates among which most are phantom words.

Nonetheless, infants preform well on various statistical learning experiments, showing the memory for statistically defined words. Why? In those experiments infants are presented with cued recall tasks, where they have to discriminate actual words from non- or part-words. This situation differs from actual recall where memory must be reconstructed from the possible syllable transitions that are stored in memory (Endress & Mehler, 2009). If infants are never asked to discriminate statistically well-defined words and phantom words, the experiments *a priori* exclude the possibility of recalling words incorrectly. Furthermore, there is only scarse experimental evidence that infants can segment continuous speech where word internal TPs are not perfect ($TP<1.0$) (Hay et al., 2011). On the one hand, infants in these experiments could assign word boundaries whenever variance occurs (e.g. when one syllable can co-occur with more than another syllable) without having to perform any statistical calculations or comparisons between word internal and word straddling syllable pairs. On the other hand, even if infants were tracking TPs, phantom words cannot emerge when word internal TPs are artificially inflated so that one syllable can only be followed by one specific syllable. We therefore continue our study by testing infants' memory for words that are statistically probable, but that were not present as words in the stream.

4.4 Study 3

Associative memory for words at 6 months

In the present study, we tested 6-month-old infants with the Head-Turn-Preference Procedure (HTPP) (Nelson et al., 1995). They were divided into 6 groups. We tested their memory for the words that they heard during the experiment (Group 1 & 2), and compared it to their memory for the phantom words – the words that have equal transitional probabilities as actual words, but cannot be found in the corpus (Groups 3–6).

4.4.1 Methods

4.4.1.1 Subjects

The final number of participants that were included in the analysis was 72 (57 boys, 58 girls, mean age 187 days (SD=17 days), all normally developing children with the APGAR scores at birth 8 or more), divided into 6 groups (N=12). From the total number of tested participants (N=115) we excluded 43 infants because they did not complete the experiments due to fuzziness. Informed consent was obtained from the parents of all infant participants before the beginning of the experiments. The infants' experiments were approved by the Bioethics Committee of SISSA – International School for Advanced Studies (date of the approval 25/03/2010). Parents received a small monetary compensation for travelling expenses.

4.4.1.2 Materials

Familiarization in all experiments was equal. Infants heard a continuous stream of words for 3 minutes, ramped at the beginning and at the end of listening, with no pauses inserted at any point in the stream. Participants heard 12 3-syllabic words (Figure 23), pseudo-randomly repeated 25-times. Each of the 9 syllables that constituted the words was 240ms long and had the monotonous peach of 200Hz. The words were synthesized with MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & Van der Vreken, 1996) using the Italian female voice IT4 (http://tcts.fpms.ac.be/synthesis/mbrola.html).

Although the word-initial and word-final syllables belonged to the same group, no word contained any syllable repetition. Adjacent transitional probability (TP) of the first syllabic pair in each word was 0.25, and 0.5 for the second pair. Because words were repeated randomly, the TPs at the word boundaries varied from 0.155 to 0.048 (see Table 10).

During the test phase, infants heard two of the possible lists of 3-syllabic items: words, part-words, non-words, or phantom words (Table 11). During each test trial, the test items were delimited by 500ms silent pauses; each item was repeated for 15 seconds. Part-words were strings that consisted of one syllable of one word and two syllables of another word. The highest possible TPs were therefore 0.5 for one syllable pair and 0.155 for the other. Non-words were novel co-occurrences of the same syllables, so that their TP in the familiarization stream was 0.0. Phantom words were composed from the same syllable pairs as words, but since each middle syllable (column B) was combined with the third syllable from another word in which the middle syllable was equal (see Figure 23), the result was a string with internal adjacent TPs as high as in the true words, but which never occurred in the familiarization stream. Due to the internal structure of the words, the second syllable in each word could either combine with the third syllable so that there was no word-internal syllable repetition (e.g. the pair ZA-GO combined with the pair GO-BI in Figure 23; ABC phantoms in Table 11), or the combination could result in the ABA structure (e. g. the pair ZA-GO being combined with the pair GO-ZA; ABA phantoms in Table 11).

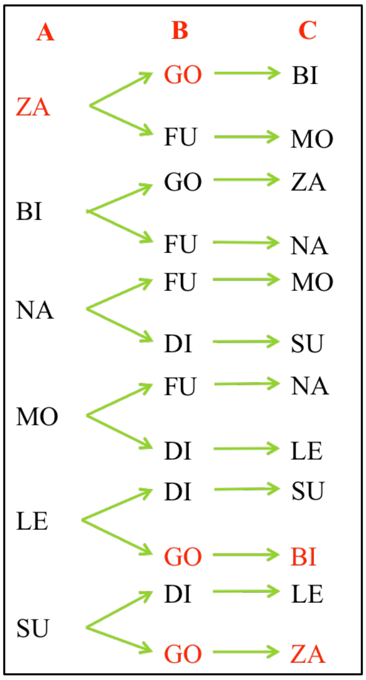

Figure 23 The list of the 12 words that were used in the familiarization and their internal syllables structure.

	BI	**DI**	**FU**	**GO**	**LE**	**MO**	**NA**	**SU**	**ZA**
BI	0	0	0.25	0.25	0.119	0.071	0.155	0.071	0.083
DI	0	0	0	0	0.5	0	0	0.5	0
FU	0	0	0	0	0	0.5	0.5	0	0
GO	0.5	0	0	0	0	0	0	0	0.5
LE	0.096	0.253	0	0.253	0	0.084	0.096	0.084	0.133
MO	0.083	0.25	0.25	0	0.143	0	0.048	0.155	0.071
NA	0.107	0.25	0.25	0	0.095	0.131	0	0.095	0.071
SU	0.119	0.25	0	0.25	0.06	0.083	0.095	0	0.143
ZA	0.095	0	0.25	0.25	0.071	0.131	0.107	0.095	0

Table 10 Adjacent transitional probabilities of all syllable pairs that occurred in the actual familiarization stream.

Words	**Part-words**	**Non-words**	**ABC phantoms**	**ABA thantoms**
zafumo	gobisu	gofule	zafuna	zagoza
bifuna	fumole	fudina	bifumo	bigobi
nadisu	disule	sufugo	nadile	nafuna
modile	bilego	digosu	modisu	mofumo
legobi	monadi	zadifu	legoza	ledile
sugoza	zalego	nagodi	sugobi	sudisu

Table 11 The list of all the items that were used during the test phases.

4.4.1.3 Procedure

Infants first heard the 3min familiarization stream of words. While listening, they saw a green ball looming randomly at one of the three 17inch screens that surrounded participants. After the offset of the familiarization, the test phase started. Each test trial was started manually after the infant's fixation to the looming ball at the central screen, followed by a fixation to the screen on the side from which the sound was to be emitted and which was indicated with a looming ball. The trial stopped when infants turned away from the screen for more than 2 seconds or after 15 seconds of uninterrupted listening and looking to the appropriate side. There were in total 12 test trials, pseudo-randomly assigned to one of the two conditions and to a side from which the sound was emitted. The stimuli were presented via PsyScope X software (http://psy.ck.sissa.it/).

Infants were seated on their parent's lap (in order to avoid any discomfort due to a new environment without a presence of a familiar person) at 65cm distance from the central and the side monitors (the distance between the side screens was 130cm, the angle between the screens was 90º). The recording booth was covered with black curtains, with only the three screens visible. To ensure that the infants' looking-behavior was not affected by parental influence, parents wore headphones and listened to classical music during the whole experiment. At the infant's seating position, the average measured volume of the auditory stimuli was 60.05dB(A) (range 56.2–63.9 dB(A) on A-scale, which represents the characteristic hearing curve of the human ear). Infants were recorded with the Sony digital camera (24 frames/sec). The recordings were analyzed frame-by-frame with the PsyCode software (http://psy.ck.sissa.it/PsyCode/PsyCode.html).

4.4.2 Results

The results are presented in Figure 24. For each group we measured the proportion of time that each of the participants spent listening to the two test lists during the 12 test trials. The average proportion of looking to a side during the presentation of the target list was then compared to the proportion of looking to a side during the presentation of the non-target list. Infants in Group 1 heard 6 words and 6 non-words during the test phase and spent on average more time looking to the words (ME=57.5%, SD=10.88; paired two-tailed t-test between the total looking times to each of the conditions: t(11)=2.371, p=0.037; Cohen's d=0.968). Infants in Group 2 heard 6 words and 6 part-words and listened longer to words again (ME=56.0%, SD=9.1; paired two-tailed t-test between the conditions: t(11)=2.286, p=0.043; Cohen's *d*=0.93). Group 3 heard ABC phantom words compared to the non-words and paid more attention to the phantom words (ME=54.4%, SD=6.1; paired two-tailed t-test between the conditions: t(11)=2.5, p=0.03; Cohen's *d*=1.021). Similarly, Group 5 listened longer to the phantom words compared to non-words even if the phantom words had an internal non-adjacent repetition regularity that was not presented in the words (ABA phantoms) (ME=57.9%, SD=7.7; paired two-tailed t-test between the conditions: t(11)=3.564, p=0.004; Cohen's *d*=1.455). On the contrary, when phantoms words were compared to the part-words – chunks of the stream that infants heard during the familiarization – infants listened longer to the part-words, both when the phantoms had a non-repetitive structure, as for Group 4 (ME=44.5%, SD=5.9; paired two-tailed t-test between the conditions: t(11)=-3.222, p=0.008; Cohen's d=1.35), and when they contained internal repetition, as for Group 6 (ME=45.1%, SD=6.4; paired two-tailed t-test between the conditions:t(11)=-2.679, p=0.021; Cohen's d=1.094). A one-way ANOVA between the 6 groups with post-hoc multiple comparisons reveals that the results in Groups 4 and 6 are significantly different from other 4 groups (F(5)=7.293, p=1.7201e-05).

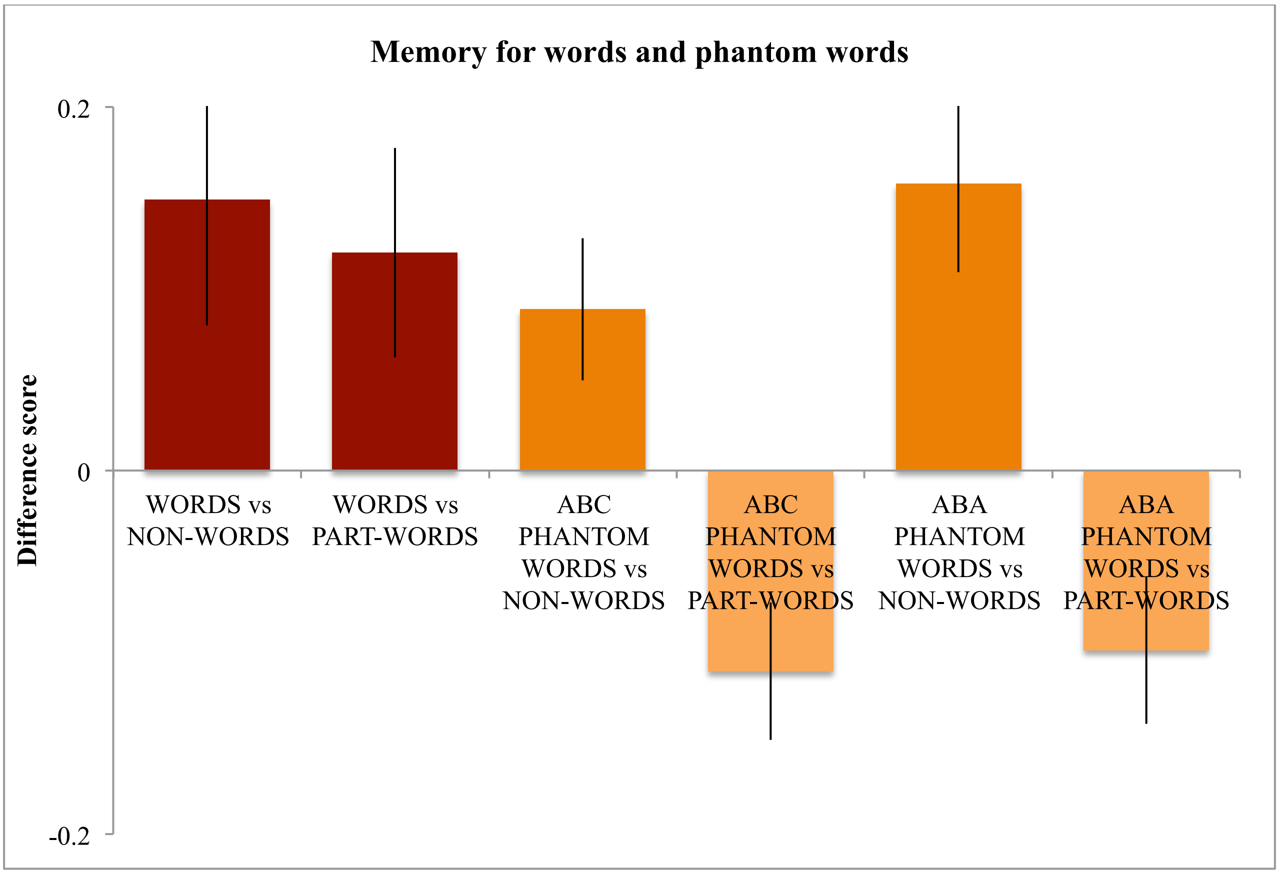

Figure 24 The results for Groups 1-6. The positive scores represent the preference for the words or phantom words, and the negative scores represent the preference for the part-words in Groups 4 and 6. The error bars represent confidence intervals.

4.4.3 Discussion

In Groups 1 and 2, infants looked longer to the familiar words when compared to the non-words and part words. Consistent difference in looking time shows that infants recognize the words as more familiar than the non-words or the part-words. Infants in these groups show familiarity preference, which is usually explained by the increased complexity of the task. The complexity in our study is increased compared to some other similar studies: we kept the same number of syllables, but we used relatively large number of words in the familiarization stream (12 words, compared to 4 in Saffran et al., 1996; Thiessen & Saffran, 2007), and the within-word TPs in the familiarization stream were lower than 1.0. For the more demanding and complex tasks, it has been often reported that infants cease to show the more frequently attested novelty preference (Gomez & Gerken, 1999; Johnson et al., 2009; Saffran et al., 1996), and show the preference for familiar items (Aslin, 2007; Gomez & Maye, 2005; Thiessen & Saffran, 2007). We conclude that infants in our experiments show looking preference for items they find more familiar because of the increased complexity of the familiarization stream.

In Groups 3 and 4, we tested whether infants exhibit some traces of false memories for statistically probable sequences that never occurred in the stream. As before, infants show familiarity preference for phantoms when they are compared to non-words. However, when they hear part-words together with phantoms, they recognize part-words and look longer to them. The same pattern of results is repeated with a slightly different version of phantom words, which introduces the repetition pattern (Groups 5 and 6). If complexity of familiarization stream induces the familiarity preference, then the preference for part-words in Groups 4 and 6 could be explained as familiarization preference as well. This can in turn lead us to explain the preference for part-words as a sign that infants treat them as more familiar compared to phantom words.

If infants indeed compute TPs in the incoming stream and recognize the words based on their relatively higher statistical regularity, why would they recognize part-words as more familiar than the phantom words that have higher word-internal TPs? One possibility is that infants in this study computed non-adjacent TPs, which were higher in part-words than in phantom words in our stream. But this would imply that they would reject the adjacent TPs that are higher than the non-adjacent ones. Whereas adults can track non-adjacent probabilities only after a long exposure (Peña et al., 2002), infants were shown to track non-adjacent probabilities only when adjacent probabilities are considerably lower than the non-adjacent and not otherwise (Gomez & Maye, 2005; Lany & Gomez, 2008; Marchetto, 2009). We therefore conclude that it is highly unlikely that infants in our study compute non-adjacent probabilities and reject the phantom words based on such computation.

Another possibility is that infants can select possible word candidates only if they receive positive evidence from the speech stream. In this case, higher frequency of occurrence should increase the familiarity of the items; and it appears that infants are sensitive to the frequency distribution in their input (Jusczyk & Hohne, 1997; Ngon et al., 2013). In our study, infants in each group recognized a minimal difference in familiarity: words are more familiar than part-words, part-words are more familiar than phantoms, and phantoms are, in turn, more familiar than non-words. Such a distribution is compliant also with the frequency measures in the familiarization stream: words were repeated 25-times, part-words occurred 6.3-times on average, and non-words are novel strings with no occurrences in the familiarization. And although phantom words were novel strings as well, their internal chunks were highly frequent (31.5 occurrences on average), creating an apparent familiarity when compared to non-words in Groups 3 and 5.[2]

We therefore conclude that our experiments do not give a positive evidence that infants memorize words based on their internal TPs. In our experiments, infants show memory traces only for the chunks that they actually heard in the stream, rejecting strings that have high adjacent probabilities but which never occurred in the familiarization stream. If we are assuming that infants reject strings that never occurred in the input, this should help them to avoid remembering numerous highly probable syllable combinations which are not real words (see 4.3).

---

[2] In another experiment (Aslin et al., 1998), infants were shown to look longer to the part-words when compared to the words, where in the familiarization stream, the frequencies were adjusted so that the part words had lower TPS, but were more frequent than the words. Based on their previous experiments, the authors interpret infants to show novelty preference and to put more weigth to the TPs than to the frequencies. Since they conducted only one experiment with the given familiarization stream, we believe the definite conclusion about the meaning of these results is not possible.

4.5

General discussion

The present chapter covers three studies that addressed the question of how infants extract words from a continuous speech stream. The possible role of statistical learning in this task was examined in more detail. In the first study, we asked whether statistical information present in natural languages vary across languages and how would possible variation across languages affect the learning curves in hypothetical learners. The results showed that in some languages, statistical regularities within words are strong enough so that the majority of words could be extracted based on co-occurrence statistics, whereas in other languages, the amount of words that could be extracted using the same algorithm was drastically reduced. If infants rely primarily on statistical regularities in earlier stages of language development (Thiessen & Saffran, 2003), cross-linguistic variation could cause a considerable delay in early word learning in some languages; yet, such a variance is not attested in language acquisition literature (Gleitman & Newport, 1995). We conclude that statistical regularities must be a language-dependent cue for word segmentation, similarly to lexical stress, phonotactics and other cues (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; E. K. Johnson, Jusczyk, Cutler, & Norris, 2003; Jusczyk et al., 1999; Nazzi, Iakimova, Bertoncini, & Alcantara, 2006). This would imply that statistical learning should be set as any other linguistic parameter, which is not a very plausible proposal. That is, statistical regularities were proposed for language acquisition to overcome or precede language-specific word segmentation (Swingley, 2005; Thiessen & Saffran, 2003).

The second study went further in making hypotheses about the memory space needed if adjacent co-occurrence regularities defined the possible words. Using the same corpora as in the first study, we showed that the proportion of statistically highly probable syllabic sequences vastly exceeds the proportion of actual words in all analyzed languages. If language learners indeed remember word candidates according to their statistical probabilities of co-occurrence, other mechanisms that help to tam associative memories must exist in these learners. Otherwise, recognizing the real words among all candidates would be an extremely difficult task. The third study brings the question of possible false memories for words in young infants. If infants really compute statistical probabilities of adjacent syllables, they might show similar memory traces for words and for statistically probable sequences of syllables that never occurred in the input. The results show that infants react to words in a different way than to phantom words. We conclude that infants rely on the positive evidence in their input, which allows them to reject phantom words as plausible word candidates.

In this study, infants actually recognize the more familiar items when they are presented in isolation during the test trials (with pauses between the words). The items that infants recognized as more familiar were not only present in the input but also more frequent than the less familiar item in each comparison. Thus, the frequency of hearing the words may have contributed to the familiarity with the words that infants exhibited during the test, which confirms other studies that point to the role of frequency in word learning (Jusczyk & Hohne, 1997; Ngon et al., 2013) Computing the co-occurrence statistics may have also contributed to the sense of familiarity with the three-syllabic words. If only adjacent transitional probabilities were computed, the sense of familiarity with phantoms should be equal to the words, and it was not. While these experiments may have answered the question of how does the input affect the sense of familiarity with particular items, they do not address the question of how infants actually learn words. The ease with which infants remember actual words suggests that statistical computations – which are present at birth (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), are not language-specific (Kirkham, Slemmer, & Johnson, 2002; Saffran et al., 1999), and which are shared with other species (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005) – must be tamed by language specific knowledge. For example, the prosody of the speech signal provides strong cues for segmenting continuous speech (Endress & Hauser, 2010; Gout, Christophe, & Morgan, 2004; Langus et al., 2012; Nespor & Vogel, 2007), it constrains statistical computations over syllables (Langus et al., 2012; Shukla et al., 2007, 2011) and eliminates the recollections of phantom words (Endress & Mehler, 2009). It is therefore possible that the human brain has evolved to use non-associative speech cues to override statistical learning, and that the speech signal has – over generations – adapted to mark boundaries in continuous speech non-statistically. This suggests that the development of higher cognitive abilities, such as human language, cannot depend solely on domain general abilities to reliably extract, encode, and recollect linguistic knowledge.

Language specific knowledge is thus not only necessary for learning about the grammar (Yang, 2004, 2013), but also for learning about the lexicon of our native language.

5

General conclusions

To this date, the debates about the essence of the human species evolved around three questions: (1) what are the cognitive capacities unique to humans and how they interact (Amati & Shallice, 2007), (2) what is their relation to low-level perceptual processes that we share with other species (Toro & Trobalón, 2005), and (3) how they develop – which of them are present in humans from birth on (or even earlier); do they maturate, or do they emerge from low-level processes. This thesis has concentrated on the third question. Our hypothesis was that infants are rational creatures that are able to make inferences about the world that surrounds them from the very beginning, adjusting these inferences in time by the accumulated experiences (Cesana-Arlotti et al., 2012; Téglás et al., 2011). Assuming this is correct, very young infants should be able to show responses that are not based on simple perceptual cues, but follow from more abstract assumptions that are not directly tied to their immediate input.

In the first part of the thesis, we show that infants modify the way they observe visual stimuli on the basis of auditory stimuli. In Experiment 3 infants switched their preference for looking at the repeated objects into the preference for looking at the novel objects exclusively on the basis of the word that they heard: if they heard the repeated words, they preferred repeated objects, and if they heard novel words, they preferred novel objects. And when the word was missing, they could not predict which object was the target, so they looked at both. Similar results are obtained in Experiment 4 where the repeated or novel items were movements. Because visual stimuli in all three conditions were equal, infants' gaze responses should be similar in all three cases if they were responding on the basis of low-level visual cues. And if the auditory stimuli were perceived just as intercensory redundancy (Bahrick & Lickliter, 2012), we would not expect sharp differences between responses to novel and familiar words, because the words were equally salient. Furthermore, if the responses were perceptually driven, when the word was missing, we would expect the same behavior as in Experiments 1 and 2 that did not contain words. We therefore conclude that the responses in Experiment 3 and 4 clearly show that infants as young as 4 months follow an abstract logic when observing the world, which goes far beyond simple perceptually based responses. The experimental design we chose resembles the situations in which language learners have to make inferences about which word refers to which object. The principle that is commonly understood to help language learners in this task is called Mutual Exclusivity and refers to the fact that in everyday situations (with some exceptions) words are used unambiguously: each word has a different referent (Halberda, 2003, 2006; Markman & Wachtel, 1988). If a novel word is encountered together with a novel referent, language learners tend to map the two. We find this principle explanatory for the results in Experiments 3 and 4. Because infants as young as 4 months exhibit such complex behavior, we might be compelled to infer that the principle guiding the responses is in some way innate in humans. As much as such inference sounds appealing, further studies are required to answer the questions of innateness and domain-specificity of early word-learning mechanisms. There is however another result that was not expected in these experiments, namely that the behavior that infants exhibited for object repetition and object labeling (Experiments 1 and 3) is matched almost completely by their behavior during movement repetition and movement labeling. Because, in the first period of language learning (before 18 months), infants on average learn many more object labels than any other labels, we predicted that their responses for the movements would be weaker than those for the objects. However, representation and labeling of the spatio-temporal information (in our case the direction or the goal of the movement) appears to be equally rapid and to follow the same logic followed for the objects.

The second part of this thesis asked further about the processes involved in learning words when multiple visual referents are possible. We explored the differences between learning auditory regularities and learning words. Infants have no trouble in learning more than one word-object pair within a single event (Smith & Yu, 2008). But when they have to learn more than one auditory regularity associated with more than one visual stimulus (reinforcer), they fail (Kovács & Mehler, 2009b). This is in contrast with the results that show that infants successfully learn multiple auditory regularities when they are not associated to any visual stimuli (Gerken et al., 2011). Our results indicated that infants fail not because the regularities themselves are too difficult to learn or to differentiate. It is therefore possible that the process of making the correct decisions about where to expect the visual reinforcers (cognitive control) itself is difficult for infants in the first year of life. In the word-learning situation, infants succeeded in learning two labels during the same event. This confirms the previous results, but this time with younger infants, and indicates that learning words requires a different level (or a different kind) of cognitive control than associating auditory regularities to visual events. We propose that the difference between learning words and learning regularities consists of the fact that two different learning processes are involved in these tasks: in the former, infants build new concepts, whereas in the latter, they arbitrarily associate an auditory and a visual input.

The third part of this thesis explored potential relationships between language-specific and domain-general cues to word segmentation. The current literature suggests that language learners might extract plausible word candidates from continuous input just by using adjacent co-occurrence regularities. Our corpus analysis showed that this appears to be possible in some languages, but less so in others. How informative these domain-general cues are, is moreover in tight connection with language-specific information about speech rhythm. Even in languages where statistical cues for word segmentation appear to be informative, infants will face the problem of syllable sequences that are very probable but do not constitute real words. Phantom syllable sequences are very numerous in all analyzed corpora, and the question is whether infants must remember all of them, if they remember word candidates based on their co-occurrence statistic. Having to remember such a large number of false candidates in order to acquire a tiny proportion of real words seems highly improbable. We therefore tested infants for their memory for such phantom word candidates. The results show than infants recognize words from the continuous stream they heard during the familiarization (that preceded the test). However, when presented with phantom words, they show greater familiarity with the part-words contained in the familiarization stream. What may contribute to the sense of familiarity with the stimuli might therefore not be the co-occurrence statistics of the adjacent syllables, but simply the frequency with which the chunks of syllables occurred in the stream. This might indicate that infants do not segment words based on their co-occurrence statistic at all, but that they require positive evidence in order for a word to be recognized as such. The present results do not answer the question of how infants actually extract words, as these results only give information about how familiar infants are with certain parts of speech. It was recently suggested that some universal prosodic cues might contribute to words segmentation (Endress & Hauser, 2010). Further studies are needed to establish whether such prosodic cues guide word segmentation also in infants.

5.1 Some open questions

Word learning appears to be a task that involves logical reasoning and cannot be explained entirely by associative processes. This holds both for the process of extracting possible word candidates from fluent speech and for the process of selecting the correct referent for a given label. Along with the conclusions we drew above, the present studies opens - or leaves open - several questions.

It is widely accepted that language is a system operating independently from the ability to reason – for example, children who grew up isolated from language show unimpaired mental abilities although their language never fully recuperated after their insertion in society (Gleitman & Newport, 1995). Similarly, pre-linguistic infants can reason about possible and probable events (Spelke & Kinzler, 2007; Téglás et al., 2011, 2007). But some parts of logic have never been shown to exist without language: it is very difficult, if not impossible, to disentangle the concept of negation from the concept of same/different (which has been attested also in other species; Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001; Pepperberg, 1999) without the usage of language. In our labeling tasks in Experiments 3 and 4, infants might have used negation to make inferences about new words in each trial. But the experiment involved labeling tasks, disentangling the concept of negation from language is not possible.

Similarly, our experiments leave open the question about categorization in word learning. Infants in Experiments 3 and 4 have reacted similarly to words that traditionally belong to the noun category and to words related to spatio-temporal relations, possibly verbs or adverbs. Across languages, infants tend to learn nouns faster than verbs, and the pace of learning has been attributed to the saliency and concreteness of nouns in comparison to verbs (Golinkoff & Hirsh-Pasek, 2008; but see Tardif, 1996). In our experiments, both objects and movements were concrete and salient, which can explain similar results in both experiments. However, these experiments leave unclear whether infants, when given a choice, attend more to the labels for objects than to the labels for the movements.

Thirdly, this thesis leaves open the question about other factors that will influence whether words will be acquired and when. According to the experiments presented in this thesis, in their first 9 months of life infants can extract words from fluent speech; they are also able to make logical inferences about the new label-referent pairs and form expectations based on the mapping between labels and objects, which are different from the expectations based on other learned cross-modal associations. Other skills required for successful word learning have been also shown to emerge in the first year of life, such as the social context of labeling (Csibra & Gergely, 2009; Csibra, 2010), the ability to categorize the visual input (Hespos & Spelke, 2004; Pruden, Roseberry, Göksun, Hirsh-Pasek, & Golinkoff, 2013; Tenenbaum et al., 2011), and the long term retention of visual and auditory stimuli (Bahrick & Pickens, 1995; Spiegel & Halberda, 2011). Nonetheless, the majority of infants' vocabulary still develops much later, during and after the second year of life. While it is well outside the scope of this thesis to speculate about the possible factors that play a role in this asymmetry, it is worth keeping in mind that we still know little about how language actually emerges in humans.

6

Bibliography

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2010). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental Science*, *2*, no–no. doi:10.1111/j.1467-7687.2010.00989.x

Amati, D., & Shallice, T. (2007). On the emergence of modern humans. *Cognition*, *103*(3), 358–85. doi:10.1016/j.cognition.2006.04.002

Anderson, R. C. (1993). *The vocabulary conundrum* (p. 22).

Antelmi, D. (2004). The Antelmi corpus. *CHILDES Database*.

Antinucci, F., & Parisi, D. (1073). Early language acquisition: A model and some data. In C. Ferguson & D. Slobin (Eds.), *Studies in Early Language Development*. Holt, New York.

Argus, R. (2004). The Argus corpus. *CHILDES Database*.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. doi:10.1111/j.1467-7687.2007.00563.x

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176. doi:10.1177/0963721412436806

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. doi:10.1111/1467-9280.00063

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). ACM press New York.

Bahrick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: the salience of actions over faces in dynamic events. *Child Development*, *73*(6), 1629–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12487483

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Brenner, D. J. Lewkowicz, & C.

Spence (Eds.), *Multisensory Development* (pp. 183–205). Oxford University Press: Oxford, Englend.

Bahrick, L. E., & Pickens, J. N. (1995). Infant memory for object motion across a period of three months: Implications for a four-phase attention function. *Journal of Experimental Child Psychology*, *59*, 343–371.

Baillargeon, R. (1987). Object Permanence in 3 1/2- and 4 1/2-Month-O1d Infants. *Developmental Psychology*, *23*(5), 655–664.

Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How Do Infants Reason About Physical Events? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development, 2nd edition* (pp. 11–48). Willey-Blackwell.

Batchelder, E. O. (2002). Bootstrapping the lexicon: a computational model of infant speech segmentation. *Cognition*, *83*(2), 167–206. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11869723

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17908–13. doi:10.1073/pnas.1205413109

Benavides-Varela, S., Macagno, F., Bion, R. A. H., Peretz, I., Gomez, D. M., & Mehler, J. (2011). Memory in the Neonate Brain. *PloS One*, *6*(11). doi:10.1371/Citation

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–8. doi:10.1073/pnas.1113380109

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: rhythmical basis of speech representations in neonates. *Language and Speech*, *38*(4), 311–329. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8816088

Bloom, P. (2001). Précis of How children learn the meanings of words. *The Behavioral and Brain Sciences*, *24*(6), 1095–103; discussion 1104–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12412326

Bloom, P. (2004). Behavior. Can a dog learn a word? *Science (New York, N.Y.)*, *304*(5677), 1605–6. doi:10.1126/science.1099899

Bol, G. W. (1995). Implicational scaling in child language acquisition: The order of production of Dutch verb constructions. In *Papers from the Dutch-German Colloquium on Language Acquisition*. Amsterdam Series in Child Language.

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294–328. doi:http://dx.doi.org/10.1016/0022-0965(83)90085-1

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, *18*(10), 924–5. doi:10.1111/j.1467-9280.2007.02002.x

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., … Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: spanish, dutch, French, hebrew, italian, korean, and american english. *Child Development*, *75*(4), 1115–39. doi:10.1111/j.1467-8624.2004.00729.x

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. doi:10.1111/j.0956-7976.2005.01531.x

Brent, M. R., & Cartwright, T. a. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1-2), 93–125. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8990969

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11376642

Carey, S. (2010). Beyond Fast Mapping. *Language Learning and Development*, *6*(3), 184–205. doi:10.1080/15475441.2010.484379.Beyond

Casefli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sander, L., & Weir, J. (1995). A Cross-Linguistic Study of Early Lexical Development. *Cognitive Development*, *10*, 159–199.

Cesana-Arlotti, N., Téglás, E., & Bonatti, L. L. (2012). The probable and the possible at 12 months: intuitive reasoning about the uncertain future. *Advances in Child Development and Behavior*, *43*, 1–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23205406

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

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*, *13*(2-3), 221–268. doi:10.1080/016909698386528

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Cloutman, L. L., Binney, R. J., Morris, D. M., Parker, G. J. M., & Lambon Ralph, M. a. (2013). Using in vivo probabilistic tractography to reveal two segregated dorsal "language-cognitive" pathways in the human brain. *Brain and Language*. doi:10.1016/j.bandl.2013.06.005

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, *25*(2), 141–168. doi:10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–53. doi:10.1016/j.tics.2009.01.005

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words* (p. 555). MIT press.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*(1/2), 103–108.

D'Anna, C., Zechmeister, E., & Hall, J. (1991). Toward a meaningful definition of vocabulary size. *Journal of Literacy Research*, *23*(1), 109–122. doi:10.1080/10862969109547729

Dehaene-Lambertz, G., Montavont, a, Jobert, a, Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53–65. doi:10.1016/j.bandl.2009.09.003

Den Os, E., & Kager, R. (1986). Extrametricality and stress in Spanish and Italian. *Lingua*, *69*, 23–48.

Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word* (p. 118). MIT Press.

Doupe, a J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631. doi:10.1146/annurev.neuro.22.1.567

Dryer, M., & Haspelmath, M. S. (2011). *The World Atlas of Language Structures Online*. (M. Dryer Matthew S. & Haspelmath, Ed.). Munich: Max Planck Digital Library. Retrieved from http://wals.info/

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vreken, O. (1996). The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. In *Proc. ICSLP'96, Philadelphia, vol. 3* (pp. 1393–1396).

Edwards, J. R. (2001). Ten Difference Score Myths. *Organizational Research Methods*, *4*(3), 265–287. doi:10.1177/109442810143005

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–99. doi:10.1016/j.cogpsych.2010.05.001

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*(3), 351–367. doi:10.1016/j.jml.2008.10.003

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348–53. doi:10.1016/j.tics.2009.05.005

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology. General*, *134*(3), 406–19. doi:10.1037/0096-3445.134.3.406

Fagan III, J. F. (1973). Infants' delayed recognition memory and forgetting. *Journal of Experimental Child Psychology*, *16*(3), 424–450. doi:http://dx.doi.org/10.1016/0022-0965(73)90005-2

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 9926–30. doi:10.1073/pnas.0709884105

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., … Reilly, J. S. (1993). *MacArthur Communicative Development Inventories (CDI)*. Paul H. Brookes Publishing.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. *Child Development*, *81*(2), 472–9. doi:10.1111/j.1467-8624.2009.01408.x

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences*, *110*(38), 15231–18424.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–6. doi:10.1073/pnas.232472899

Fourtassi, A., Orschinger, B. B., Dupoux, E., & Johnson, M. (2013). Whyisenglishsoeasytosegment? In *CMCL 2013 Cognitive Modeling and Computational Linguistics Proceedings of the Workshop* (pp. 1–10). The Association for Computational Linguistics.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–25. doi:10.1016/j.cognition.2010.07.005

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: simulations and data. *Journal of Experimental Psychology. General*, *133*(3), 382–97. doi:10.1037/0096-3445.133.3.382

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology : CB*, *17*(14), 1208–11. doi:10.1016/j.cub.2007.06.011

Gambell, T., & Yang, C. (2006). Word segmentation: Quick but not dirty. *Unpublished Manuscript, Available At< Http://www. Ling. Upenn. Edu/ycharles/papers/quick. Pdf.*

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486–90. doi:10.3758/PBR.16.3.486

Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. (Ed) Guo, E. (Ed) Lieven, N. (Ed) Budwig, S. (Ed) Ervin-Tripp, K. (Ed) Nakamura, & Ş. (Ed) Õzçalişkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480). New York, NY, US: Psychology Press.

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid "labouring in vain" by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–9. doi:10.1111/j.1467-7687.2011.01046.x

Gerken, L., & Bollt, A. (2008). Three Exemplars Allow at Least Some Linguistic Generalizations: Implications for Generalization Mechanisms and Constraints. *Language Learning and Development*, *4*(3), 228–248. doi:10.1080/15475440802143117

Gervain, J., & Guevara Erra, R. (2012). The statistical signature of morphosyntax: a study of Hungarian and Italian infant-directed speech. *Cognition*, *125*(2), 263–87. doi:10.1016/j.cognition.2012.06.010

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14222–7. doi:10.1073/pnas.0806530105

Gibson, E. J., Owsley, C. J., & Johnston, J. (1978). Perception of invariants by five-month-old infants: Differentiation of two types of motion. *Developmental Psychology*, *14*(4), 407–415. doi:10.1037/0012-1649.14.4.407

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. (2001). The concepts of "sameness" and "difference" in an insect. *Nature*, *410*(April), 930–933.

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

Gleitman, L. R., & Newport, E. L. (1995). The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language. In L. R. Gleitman & M. Liberman (Eds.), *Language: An Invitation to Cognitive Science (2nd edition)*. Cambridge, MA: The MIT Press.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, *20*(3), 347–7976.

Goga, I. (2006). *Educating attention in early development of imitation, language and object manipulation abilities in human infants*. University Babes-Bolyai, Faculty of Psychology and Education Sciences, Cluj-Napoca.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–49. doi:10.1006/jecp.1998.2438

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory Redundancy and 7-Month-Old Infants' Memory for Arbitrary Syllable-Object Relations. *Infancy*, *2*(2), 219–231. doi:10.1207/S15327078IN0202_7

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to Maternal Multimodal Naming by 6- to 8-Month-Old Infants and Learning of Word-Object Relations. *Infancy*, *9*(3), 259–288. doi:10.1207/s15327078in0903_1

Goldwater, S., Griffiths, T. L., & Johnson, M. (2007). Distributional cues to word boundaries: Context is important. In *Proceedings of the 31st annual Boston University conference on language development*. Cascadilla Press.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: exploring the effects of context. *Cognition*, *112*(1), 21–54. doi:10.1016/j.cognition.2009.03.008

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, *12*(10), 397–403. doi:10.1016/j.tics.2008.07.003

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 9–125.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10349760

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 77–81.

Gomez, R., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy*, *7*(2), 183–206.

Goodale, M. A., & Milner, D. A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1374953

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567. doi:10.1016/j.jml.2004.07.002

Gredebäck, G., & von Hofsten, C. (2004). Infants' Evolving Representations of Object Motion During Occlusion: A Longitudinal Study of 6- to 12-Month-Old Infants. *Infancy*, *6*(2), 165–184.

Gredebäck, G., & von Hofsten, C. (2007). Taking an action perspective on infant's object representations. *Progress in Brain Research*, *164*, 265–82. doi:10.1016/S0079-6123(07)64015-1

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–B34. doi:10.1016/S0

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–44. doi:10.1016/j.cogpsych.2006.04.003

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*(2), 190–222.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11124355

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. doi:10.1016/j.cogpsych.2011.06.002.Linking

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. *Cognition and the Development of Language*, 221–234.

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*(22 July), 453–456.

Hochmann, J.-R. (2013). Word frequency, function words and the second gavagai problem. *Cognition*, *128*(1), 13–25. doi:10.1016/j.cognition.2013.02.014

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental Science*, *14*(6), 1445–58. doi:10.1111/j.1467-7687.2011.01089.x

Ichinco, D., Frank, M. C., & Saxe, R. (2009). Cross-situational Word Learning Respects Mutual Exclusivity. In *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 2214–2219).

Ishii, T. (1999). The JUN corpus.

Jackson-Maldonado, D., & Thal, D. (1994). Lenguaje y cognición en los primeros años de vida: resultados preliminares. *Rev. Psicología Y Sociedad*, *25*, 21–27.

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*(4), 548–567. doi:10.1006/jmla.2000.2755

Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, *46*(1), 65–285.

Johnson, S. P., Amso, D., & Slemmer, J. a. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(18), 10568–73. doi:10.1073/pnas.1630655100

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*(1), 94–108. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16150000

Johnson, S. P., Fernandas, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. a. (2009). Abstract Rule Learning for Visual Sequences in 8- and 11-Month-Olds. *Infancy : The Official Journal of the International Society on Infant Studies*, *14*(1), 2–18. doi:10.1080/15250000802569611

Jusczyk, P. W., Cutler, a, & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8339688

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' Memory for Spoken Words. *Science (New York, N.Y.)*, *277*(5334), 1984–1986. doi:10.1126/science.277.5334.1984

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, *39*(3-4), 159–207. doi:10.1006/cogp.1999.0716

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, *21*(1), 50–54. doi:10.3758/BF03199467

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". *Science (New York, N.Y.)*, *304*(5677), 1682–3. doi:10.1126/science.1097859

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Kohler, K. (2004). The Kohler corpus. *CHILDES Database*.

Korgesaar, H. (2011). The Korgesaar corpus. *CHILDES Database*.

Korman, M. (1984). Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language*, *5*, 44–55.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., … Dehaene-Lambertz, G. (2013). A Neural Marker of Perceptual Consciousness in Infants. *Science (New York, N.Y.)*, *340*(6130), 376–380. doi:10.1126/science.1232509

Kovács, A. M. (2008). *Learning two languages simultaneously: mechanisms recruited for dealing with a mixed linguistic input*.

Kovács, A. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6556–60. doi:10.1073/pnas.0811323106

Kovács, A. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science (New York, N.Y.)*, *325*(5940), 611–2. doi:10.1126/science.1173947

Kovács, A. M., & Mehler, J. (2009c). Flexible learning of multiple speech structures in bilingual infants - Supporting material. *Science (New York, N.Y.)*, *325*(5940), 611–2. doi:10.1126/science.1173947

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews. Neuroscience*, *12*(4), 217–30. doi:10.1038/nrn3008

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–43. doi:10.1038/nrn1533

Landau, B., & Gleitman, L. R. (2009). *Language and Experience: Evidence from the Blind Child* (p. 250). Harvard University Press.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*(1), 285–306. doi:10.1016/j.jml.2011.09.004

Lany, J., & Gomez, R. L. (2008). Twelve-Month-Old Infants Benefit From Prior Experience in Statistical Learning. *Psychological Science*, *19*(12), 1247–1252. doi:10.1111/j.1467-9280.2008.02233.x.Twelve-Month-Old

Leslie, a M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: developing `what' and `where' systems. *Trends in Cognitive Sciences*, *2*(1), 10–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21244957

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition*, *89*(3), 295–303. doi:10.1016/S0010-0277(03)00116-1

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, *58*(4), 441–86. doi:10.1016/j.cogpsych.2008.11.001

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff Michnick, R. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. Golinkoff Michnick (Eds.), *Action meets word: How children Learn verbs* (pp. 364–392).

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Haryu, E., Vanegas, S., … Sanchez-Davis, B. (2010). A developmental shift from similar to language-specific strategies in verb acquisition: a comparison of English, Spanish, and Japanese. *Cognition*, *114*(3), 299–319. doi:10.1016/j.cognition.2009.10.002

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397–405. doi:10.1016/j.neuron.2009.07.012

Makhoul, J., Kubala, F., Schwartz, R., & Weischedel, R. (1999). Performance measures for information extraction. In *Proceedings of DARPA Broadcast News Workshop* (pp. 249–252).

Marchetto, E. (2009). *Discovering words and rules from speech input: an investigation into early morphosyntactic acquisition mechanisms*.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science (New York, N.Y.)*, *283*(5398), 77–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9872745

Markman, E. M., & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, *8*(11), 479–81; discussion 481. doi:10.1016/j.tics.2004.09.007

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–285.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275. doi:10.1016/S0010-0285(03)00034-3

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society B*, *200*, 269–294.

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157–77. doi:10.1111/j.1551-6709.2012.01239.x

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655. doi:10.1037//0096-1523.27.3.644

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–94. doi:10.1006/cogp.1999.0721

Matuz, T., Govindan, R. B., Preissl, H., Siegel, E. R., Muenssinger, J., Murphy, P., … Eswaran, H. (2012). Habituation of visual evoked responses in neonates and fetuses: a MEG study. *Developmental Cognitive Neuroscience*, *2*(3), 303–16. doi:10.1016/j.dcn.2012.03.001

McMurray, B., & Aslin, R. N. (2004). Anticipatory Eye Movements Reveal Infants' Auditory and Visual Categories. *Infancy*, *6*(2), 203–229. doi:10.1207/s15327078in0602_4

Mehler, J. (1981). Syllables as Units in InfantSpeech Perception *. *Infant Behavior and Development*, *260*(4), 247–260.

Mehler, J., & Nespor, M. (2004). Linguistic rhythm and the acquisition of language. In A. Belletti & L. Rizzi (Eds.), *Structures and Beyond: The Cartography of Syntactic Structures* (pp. 213–221). Oxford: Oxford University Press.

Mehler, J., Nespor, M., & Peña, M. (2008). What Infants Know and What They have to Learn about Language. *European Review*, *16*(04), 429. doi:10.1017/S1062798708000392

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*(6), 513–3932.

Millotte, S., Morgan, J., Margules, S., Bernal, S., Dutat, M., & Christophe, A. (2010). Phrasal prosody constraints word segmentation in French 16-month-olds. *Journal of Portugese Lunguistics*.

Möhring, W., & Frick, A. (2013). Touching up mental rotation: effects of manual experience on 6-month-old infants' mental object rotation. *Child Development*, *84*(5), 1554–65. doi:10.1111/cdev.12065

Moore, D. S., & Johnson, S. P. (2008). Mental Rotation in Human Infants: A Sex Difference. *Psychological Science*, *19*(11), 1063–1066. doi:10.1111/j.1467-9280.2008.02200.x.Mental

Moore, D. S., & Johnson, S. P. (2011). Mental Rotation of Dynamic, Three-Dimensional Stimuli by 3-Month-Old Infants. *Infancy*, *16*(4), 435–445. doi:10.1111/j.1532-7078.2010.00058.x

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(39), 15953–8. doi:10.1073/pnas.1204319109

Narasimhan, R. (2004). The Tamil coprus. *CHILDES Database*.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9627414

Nazzi, T., Iakimova, G., Bertoncini, J., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*, 283–299. doi:10.1016/j.jml.2005.10.004

Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The Head-Turn Preference Procedure for Testing Auditory Perception. *Infant Behavior and Development*, *18*, 111–116.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In *The Blackwell companion to phonology 5* (Vol. 5, pp. 1147–1159).

Nespor, M., & Vogel, I. (2007). *Prosodic phonology: with a new foreword* (Vol. 28 %@ 3110). Walter de Gruyter.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Developmental Science*, *16*(1), 24–34. doi:10.1111/j.1467-7687.2012.01189.x

Öhman, A. (1997). As fast as the blink of an eye: Evolutionary preparedness for preattentive processing of threat. In P. J. Lang, R. F. Simons, & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 165–184). Lawrence Erlbaum Associates, Inc.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–84. doi:10.1016/j.cognition.2012.10.008

Oshima-Takane, Y. MacWhinney, B., Sirai, H., Miyata, S., & Naka, N. (1998). *CHILDES for Japanese*. The JCHAT Project Nagoya: Chukyo University.

Ota, M. (2003). *The Development of Prosodic Structure in Early Words: Continuity, Divergence and Change*. John Benjamins, Amsterdam.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–85. doi:10.1111/j.1467-8624.2009.01290.x

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science (New York, N.Y.)*, *298*(5593), 604–7. doi:10.1126/science.1072901

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705.

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3823–8. doi:10.1073/pnas.0914326107

Pepperberg, I. M. (1999). *The Alex studies: cognitive and communicative abilities of grey parrots* (p. 488). Cambridge, MA: Harvard University Press.

Piaget, J., & Cook, M. T. (1952). *The origins of intelligence in children* (p. 419). New York, NY, US: W W Norton & Co.

Pinker, S. (1994). *The language instinct* (p. 435). Penguin Books.

Pruden, S. M., Hirsh-pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words : Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development*, *77*(2), 266–280.

Pruden, S. M., Roseberry, S., Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Infant categorization of path relations during dynamic events. *Child Development*, *84*(1), 331–45. doi:10.1111/j.1467-8624.2012.01843.x

Quine, W. V. (2013). *Word and object*. MIT press.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463–475.

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*(1), 96–121. doi:10.1016/j.cogpsych.2009.02.002

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes : perceptual studies. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1–6). ICPhS, Universitat Autònoma de Barcelona.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science (New York, N.Y.)*, *288*(5464), 349–351. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10764650

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), 265–292. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10908711

Roediger, H. L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words Not Presented in Lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(4), 803–814.

Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. doi:10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928.

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10193055

Sebastián-Gallés, N. (2006). Native-language sensitivities: evolution in the first year of life. *Trends in Cognitive Sciences*, *10*(6), 239–41. doi:10.1016/j.tics.2006.04.009

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. doi:10.1016/j.cogpsych.2006.04.002

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–43. doi:10.1073/pnas.1017617108

Skoruppa, K., Cristià, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *The Journal of the Acoustical Society of America*, *130*(1), EL50–5. doi:10.1121/1.3590169

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Smoczynska, M. (1985). The acquisition of Polish. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (pp. 595–686). Lawrence Erlbaum Associates, Hillsdale, NJ.

Spelke, E. S. (1983). Constraints on the development of intermodal perception. In L. S. Liben (Ed.), *Piaget and the foundations of knowledge* (pp. 43–48). Hillsdale, NJ: Erlbaum.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 277–311). Th MIT Press.

Spelke, E. S., Breinlinger, K., Jacobson, K., & Phillips, A. (1993). Gestalt relations and object perception: a developmental study. *Perception*, *22*(12), 1483–501. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8090623

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., & Breinlinger, K. (1994). Early knowledge of object motion: continuity and inertia. *Cognition*, *51*(2), 131–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8168357

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. doi:10.1111/j.1467-7687.2007.00569.x

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–40. doi:10.1016/j.jecp.2010.10.013

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. doi:10.1016/j.cogpsych.2004.06.001

Swingley, D. (2010). Fast Mapping and Slow Mapping in Children's Word Learning. *Language Learning and Development*, *6*, 179–183. doi:10.1080/15475441.2010.484412

Tahakashi, E., & Lidz, J. (2007). Beyond Statistical Learning in Syntax. In *Proceedings of Generative Approaches to Language Acquisition* (pp. 1–10).

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492–504. doi:0.1037/0012-1649.32.3.492

Tarr, M. J., & Pinker, S. (1989). Mental Rotation and Orientation-Dependence Shape Recognition. *Cognitive Psychology*, *282*(21), 233–282.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*(48), 19156–19159.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science (New York, N.Y.)*, *332*(6033), 1054–9. doi:10.1126/science.1196404

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21. doi:10.1186/1471-2202-10-21

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science (New York, N.Y.)*, *331*(6022), 1279–85. doi:10.1126/science.1192788

Thiessen, E. D., Hill, E. a., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. doi:10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. doi:10.1037/0012-1649.39.4.706

Thiessen, E. D., & Saffran, J. R. (2007). Learning to Learn: Infants' Acquisition of Stress-Based Strategies for Word Segmentation. *Language Learning and Development*, *3*(1), 73–100. doi:10.1207/s15473341lld0301_3

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Tonelli, L. (2004). The Tonelli corpus. *CHILDES Database*.

Topál, J., Gergely, G., Miklósi, A., Erdohegyi, A., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science (New York, N.Y.)*, *321*(5897), 1831–4. doi:10.1126/science.1161437

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–56. doi:10.1016/j.cogpsych.2012.10.001

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–76. doi:10.1121/1.3129127

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.

Van de Weijer, J. (1998). *Language input for word discovery*. Unpublished doctoral dissertation, MPI Series in Psycholinguistics 9.

Van Giffen, K., & Haith, M. M. (1984). Infant visual response to Gestalt geometric forms. *Infant Behavior and Development*, *7*(3), 335–346.

Van Kampen, N. J. (1994). The learnability of the left branch condition. *Linguistics in the Netherlands*, *199*, 83–94.

Vija, M. (2004). The Vija corpus. *CHILDES Database*.

Vila, I. (1990). *Adquisición y desarrollo del lenguaje*. Graó, Barcelona.

Volterra, V. (1976). A few remarks on the use of the past participle in child language. *Journal of Italian Linguistics*, *2*, 149–157.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal : the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–64. doi:10.1111/j.1467-7687.2007.00549.x

Wang, Q., Bolhuis, J., Rothkopf, C. a, Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PloS One*, *7*(2), e30884. doi:10.1371/journal.pone.0030884

Waxman, S., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: new evidence from 11-month-olds. *Developmental Science*, *6*(2), 128–135. doi:10.1111/1467-7687.00262

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, (September). doi:10.1177/0956797613488145

Weist, R., & Witkowska-Stadnik, K. (1986). Basic relations in child language and the word order myth. *International Journal of Psychology*, *21*, 363–381.

Wexler, K. (2013). Lenneberg's dream: Learning, normal language development, and specific language impairment. In Y. Levy & J. C. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment*. Psychology Press.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language*, *31*(6), 374–255. doi:http://dx.doi.org/10.1016/0749-596X(92)90037-X.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*(2), 111–53. doi:10.1006/cogp.1996.0005

Yang, C. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451–6. doi:10.1016/j.tics.2004.08.006

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6324. doi:10.1073/pnas.1216803110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1216803110

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. doi:10.1006/jmla.2002.2864

Yu, C., & Smith, L. B. (2011). What You Learn is What You See : Using Eye Movements to Study Infant Cross-Situational Word Learning. *Developmental Science*, *14*(2), 165–180.

Yurovsky, D., Hidaka, S., & Wu, R. (2012). Quantitative linking hypotheses for infant eye movements. *PloS One*, *7*(10), e47419. doi:10.1371/journal.pone.0047419

Appendix 1

Segmentation algorithms and evaluation (python script)

```
#!/usr/bin/python
import collections
import fileinput
import sys
import numpy
import math
from scipy import stats

class Counter(collections.Counter):
    def __str__(self):
        return "\n".join("{}\t{}".format("-".join(key) if isinstance(key, tuple) else key, value)
                    for key, value in self.items())

#print "Reading %s..." %sys.argv[1]

for line in fileinput.input(sys.argv[1]):
        words = [word.split("-") for word in line.split()]

words_all = [word for word in line.split()]
freq_wordsall = Counter(words_all)
syllables = [syllable for word in words for syllable in word]
freq_syll = Counter(syllables)

bigrams_all = zip(syllables[0:-1],syllables[1:])
freq_bigrams_all = Counter(bigrams_all)
tp_bigrams_all = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_all.items())
btp_bigrams_all = dict((bigram,float(freq)/freq_syll[bigram[1]]) for bigram,freq in
freq_bigrams_all.items())
mi_bigrams_all = dict( (bigram, math.log((float(freq)/(freq_syll[bigram[0]]*freq_syll[bigram[1]])),2) )
for bigram,freq in freq_bigrams_all.items())

bigrams_int = [bigram for word in words for bigram in zip(word[0:-1],word[1:])]
freq_bigrams_int = Counter(bigrams_int)
tp_bigrams_int = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_int.items())
btp_bigrams_int = dict((bigram,float(freq)/freq_syll[bigram[1]]) for bigram,freq in
freq_bigrams_int.items())
mi_bigrams_int = dict( (bigram, math.log((float(freq)/(freq_syll[bigram[0]]*freq_syll[bigram[1]])),2) )
for bigram,freq in freq_bigrams_int.items())

bigrams_acr = [(word1[-1], word2[0]) for (word1, word2) in zip(words[0:-1], words[1:])]
freq_bigrams_acr = Counter(bigrams_acr)
tp_bigrams_acr = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_acr.items())
btp_bigrams_acr = dict((bigram,float(freq)/freq_syll[bigram[1]]) for bigram,freq in
freq_bigrams_acr.items())
mi_bigrams_acr = dict( (bigram, math.log((float(freq)/(freq_syll[bigram[0]]*freq_syll[bigram[1]])),2)
) for bigram,freq in freq_bigrams_acr.items())
```

```
TPi = sum(tp_bigrams_int.values())/len(tp_bigrams_int) if len(tp_bigrams_int)!=0 else 0
BTPi = sum(btp_bigrams_int.values())/len(btp_bigrams_int) if len(btp_bigrams_int)!=0 else 0
MIi = sum(mi_bigrams_int.values())/len(mi_bigrams_int) if len(mi_bigrams_int)!=0 else 0

TPa = sum(tp_bigrams_acr.values())/len(tp_bigrams_acr) if len(tp_bigrams_acr)!=0 else 0
BTPa = sum(btp_bigrams_acr.values())/len(btp_bigrams_acr) if len(btp_bigrams_acr)!=0 else 0
MIa = sum(mi_bigrams_acr.values())/len(mi_bigrams_acr) if len(mi_bigrams_acr)!=0 else 0



"""
TPdif = TPi - TPa
BTPdif = BTPi - BTPa
MIdif = MIi - MIa

print MIi
#print min(btp_bigrams_int.values()), max(btp_bigrams_int.values())
print MIa
#print min(btp_bigrams_acr.values()), max(btp_bigrams_acr.values())
print MIdif
print

sys.exit()
"""


###### TPs
#input
for line in fileinput.input(sys.argv[2]):
    syls = [syl for syl in line.split()]

# absolute threshold (Absolute algorithm)
cwordsTP=[]
cwords = []
last_syl = syls[0]
last_word = [last_syl]
cwords.append(last_word)
for syl in syls[1:]:
        if tp_bigrams_all[last_syl,syl] <= TPi or last_syl=="UB" or syl=="UB":
                last_word = []
                cwords.append(last_word)
        last_word.append(syl)
        last_syl=syl
cwordsTP.append(cwords)
#local minima (Relative algorithm)
lwords=[]
prelast=syls[0]
last=syls[1]
syl=syls[2]
lword=[prelast,last]
lwords.append(lword)
for next in syls[3:]:
        if tp_bigrams_all[prelast,last] > tp_bigrams_all[last,syl] < tp_bigrams_all[syl,next] or
last=="UB" or syl=="UB":
                lword = []
                lwords.append(lword)
        lword.append(syl)
        prelast=last
```

```
        last=syl
        syl=next
lwords.append([next])
cwordsTP.append(lwords)

#evaluation:
hit = 0
hitTP = []
fa = 0
faTP = []
miss = 0
missTP = []
cr = 0
crTP = []
for e in range (0, 2):
    hit, fa, miss = 0, 0, 0
    adv_wi, adv_cwi = 0, 0
    wi, cwi = 0, 0
    while wi+adv_wi < len(words) and cwi+adv_cwi < len(cwordsTP[e]):
        if not adv_wi and not adv_cwi:
            wn, cwn = len(words[wi]), len(cwordsTP[e][cwi])
        elif adv_wi:
            wi += 1
            wn += len(words[wi])
        elif adv_cwi:
            cwi += 1
            cwn += len(cwordsTP[e][cwi])
        else:
            raise RuntimeError
        if wn == cwn:
            if not adv_wi and not adv_cwi:
                hit +=1
            wi += 1
            cwi += 1
            adv_wi, adv_cwi = 0, 0
        else:
            if wn < cwn:
                adv_wi = 1
                adv_cwi = 0
            else:
                adv_wi = 0
                adv_cwi = 1
    fa = len(cwordsTP[e]) - hit
    miss = len(words) - hit
    faTP.append(fa)
    hitTP.append(hit)
    missTP.append(miss)
    cr = hitTP[e]
    crTP.append(cr)

#computing F-score
hrTP = []
frTP = []
cmpTP = []
fscoreTP = []
aprimeTP = []
for g in range (0, 2):
    hr = float(hitTP[g])/(hitTP[g] + missTP[g]) if (hitTP[g]+missTP[g])!=0 else 0
```

```
        hrTP.append(hr)
        fr = float(faTP[g])/(faTP[g] + crTP[g]) if (faTP[g]+crTP[g])!=0 else 0
        frTP.append(fr)
        cmp = float(hitTP[g])/(hitTP[g] + faTP[g]) if (hitTP[g]+faTP[g])!=0 else 0
        cmpTP.append(cmp)
        if (hrTP[g] > frTP[g]):
                aprime = 0.5 + (((hrTP[g]-frTP[g])*(1+hrTP[g]-frTP[g]))/(4*hrTP[g]*(1-frTP[g]))) if
(4*hrTP[g]*(1-frTP[g]))!=0 else 0
        else:
                aprime = 0.5 - (((frTP[g]-hrTP[g])*(1+frTP[g]-hrTP[g]))/(4*frTP[g]*(1-hrTP[g]))) if
(4*frTP[g]*(1-hrTP[g]))!=0 else 0
        aprimeTP.append(aprime)
        fscore = float(2*hrTP[g]*cmpTP[g]) / (hrTP[g] + cmpTP[g]) if (hrTP[g]+cmpTP[g])!=0 else 0
        fscoreTP.append(fscore)

#print hrTP
#print frTP
print fscoreTP
print


#### BTPs

#input
for line in fileinput.input(sys.argv[2]):
    syls = [syl for syl in line.split()]


cwordsBTP=[]
cwords = []
last_syl = syls[0]
last_word = [last_syl]
cwords.append(last_word)
for syl in syls[1:]:
        if btp_bigrams_all[last_syl,syl] <= BTPi or last_syl=="UB" or syl=="UB":
                last_word = []
                cwords.append(last_word)
        last_word.append(syl)
        last_syl=syl
cwordsBTP.append(cwords)
#local minima
lwords=[]
prelast=syls[0]
last=syls[1]
syl=syls[2]
lword=[prelast,last]
lwords.append(lword)
for next in syls[3:]:
        if btp_bigrams_all[prelast,last] > btp_bigrams_all[last,syl] < btp_bigrams_all[syl,next] or
last=="UB" or syl=="UB":
                lword = []
                lwords.append(lword)
        lword.append(syl)
        prelast=last
        last=syl
        syl=next
lwords.append([next])
cwordsBTP.append(lwords)
```

```
#evaluation:
hit = 0
hitBTP = []
fa = 0
faBTP = []
miss = 0
missBTP = []
cr = 0
crBTP = []
for e in range (0, 2):
	hit, fa, miss = 0, 0, 0
	adv_wi, adv_cwi = 0, 0
	wi, cwi = 0, 0
	while wi+adv_wi < len(words) and cwi+adv_cwi < len(cwordsBTP[e]):
		if not adv_wi and not adv_cwi:
			wn, cwn = len(words[wi]), len(cwordsBTP[e][cwi])
		elif adv_wi:
			wi += 1
			wn += len(words[wi])
		elif adv_cwi:
			cwi += 1
			cwn += len(cwordsBTP[e][cwi])
		else:
			raise RuntimeError
		if wn == cwn:
			if not adv_wi and not adv_cwi:
				hit +=1
			wi += 1
			cwi += 1
			adv_wi, adv_cwi = 0, 0
		else:
			if wn < cwn:
				adv_wi = 1
				adv_cwi = 0
			else:
				adv_wi = 0
				adv_cwi = 1
	fa = len(cwordsBTP[e]) - hit
	miss = len(words) - hit
	faBTP.append(fa)
	hitBTP.append(hit)
	missBTP.append(miss)
	cr = hitBTP[e]
	crBTP.append(cr)


hrBTP = []
frBTP = []
cmpBTP = []
fscoreBTP = []
aprimeBTP = []
for g in range (0, 2):
	hr = float(hitBTP[g])/(hitBTP[g] + missBTP[g]) if (hitBTP[g]+missBTP[g])!=0 else 0
	hrBTP.append(hr)
	fr = float(faBTP[g])/(faBTP[g] + crBTP[g]) if (faBTP[g]+crBTP[g])!=0 else 0
	frBTP.append(fr)
	cmp = float(hitBTP[g])/(hitBTP[g] + faBTP[g]) if (hitBTP[g]+faBTP[g])!=0 else 0
```

```
        cmpBTP.append(cmp)
        if (hrBTP[g] > frBTP[g]):
                aprime = 0.5 + (((hrBTP[g]-frBTP[g])*(1+hrBTP[g]-frBTP[g]))/(4*hrBTP[g]*(1-
frBTP[g]))) if (4*hrBTP[g]*(1-frBTP[g]))!=0 else 0
        else:
                aprime = 0.5 - (((frBTP[g]-hrBTP[g])*(1+frBTP[g]-hrBTP[g]))/(4*frBTP[g]*(1-
hrBTP[g]))) if (4*frBTP[g]*(1-hrBTP[g]))!=0 else 0
        aprimeBTP.append(aprime)
        fscore = float(2*hrBTP[g]*cmpBTP[g]) / (hrBTP[g] + cmpBTP[g]) if
(hrBTP[g]+cmpBTP[g])!=0 else 0
        fscoreBTP.append(fscore)

#print hrBTP
#print frBTP
print fscoreBTP
print

#### MI

#input
for line in fileinput.input(sys.argv[2]):
    syls = [syl for syl in line.split()]



cwordsMI=[]
cwords = []
last_syl = syls[0]
last_word = [last_syl]
cwords.append(last_word)
for syl in syls[1:]:
        if mi_bigrams_all[last_syl,syl] <= MIi or last_syl=="UB" or syl=="UB":
                last_word = []
                cwords.append(last_word)
        last_word.append(syl)
        last_syl=syl
cwordsMI.append(cwords)
lwords=[]
prelast=syls[0]
last=syls[1]
syl=syls[2]
lword=[prelast,last]
lwords.append(lword)
for next in syls[3:]:
        if mi_bigrams_all[prelast,last] > mi_bigrams_all[last,syl] < mi_bigrams_all[syl,next] or
last=="UB" or syl=="UB":
                lword = []
                lwords.append(lword)
        lword.append(syl)
        prelast=last
        last=syl
        syl=next
lwords.append([next])
cwordsMI.append(lwords)


#evaluation:
hit = 0
```

```
hitMI = []
fa = 0
faMI = []
miss = 0
missMI = []
cr = 0
crMI = []
for e in range (0, 2):
	hit, fa, miss = 0, 0, 0
	adv_wi, adv_cwi = 0, 0
	wi, cwi = 0, 0
	while wi+adv_wi < len(words) and cwi+adv_cwi < len(cwordsMI[e]):
		if not adv_wi and not adv_cwi:
			wn, cwn = len(words[wi]), len(cwordsMI[e][cwi])
		elif adv_wi:
			wi += 1
			wn += len(words[wi])
		elif adv_cwi:
			cwi += 1
			cwn += len(cwordsMI[e][cwi])
		else:
			raise RuntimeError
		if wn == cwn:
			if not adv_wi and not adv_cwi:
				hit +=1
			wi += 1
			cwi += 1
			adv_wi, adv_cwi = 0, 0
		else:
			if wn < cwn:
				adv_wi = 1
				adv_cwi = 0
			else:
				adv_wi = 0
				adv_cwi = 1
	fa = len(cwordsMI[e]) - hit
	miss = len(words) - hit
	faMI.append(fa)
	hitMI.append(hit)
	missMI.append(miss)
	cr = hitMI[e]
	crMI.append(cr)


hrMI = []
frMI = []
cmpMI = []
fscoreMI = []
aprimeMI = []
for g in range (0, 2):
	hr = float(hitMI[g])/(hitMI[g] + missMI[g]) if (hitMI[g]+missMI[g])!=0 else 0
	hrMI.append(hr)
	fr = float(faMI[g])/(faMI[g] + crMI[g]) if (faMI[g]+crMI[g])!=0 else 0
	frMI.append(fr)
	cmp = float(hitMI[g])/(hitMI[g] + faMI[g]) if (hitMI[g]+faMI[g])!=0 else 0
	cmpMI.append(cmp)
	if (hrMI[g] > frMI[g]):
		aprime = 0.5 + (((hrMI[g]-frMI[g])*(1+hrMI[g]-frMI[g]))/(4*hrMI[g]*(1-frMI[g]))) if
```

```
(4*hrMI[g]*(1-frMI[g]))!=0 else 0
        else:
                aprime = 0.5 - (((frMI[g]-hrMI[g])*(1+frMI[g]-hrMI[g]))/(4*frMI[g]*(1-hrMI[g]))) if
(4*frMI[g]*(1-hrMI[g]))!=0 else 0
        aprimeMI.append(aprime)
        fscore = float(2*hrMI[g]*cmpMI[g]) / (hrMI[g] + cmpMI[g]) if (hrMI[g]+cmpMI[g])!=0 else 0
        fscoreMI.append(fscore)



#print hrMI
#print frMI
print fscoreMI
```

Appendix 2

Finding phantom words in the corpus (python script)

```
#!/usr/bin/python
import collections
import fileinput
import sys
import sets
import time

start = time.time()

class Counter(collections.Counter):
    def __str__(self):
        return "\n".join("{}\t{}".format("-".join(key) if isinstance(key, tuple) else key, value)
                    for key, value in self.items())

print "Reading %s..." %sys.argv[1]

words=[]
for line in fileinput.input():
        words.extend ([word.split("-") for word in line.split()])
        words2 = [word for word in words if len(word)==2]
        words3 = [word for word in words if len(word)==3]
        words4 = [word for word in words if len(word)==4]
        words5 = [word for word in words if len(word)==5]


        syllables = [syllable for word in words for syllable in word]
        freq_syll = Counter(syllables)

        #-----

        bigrams_all = zip(syllables[0:-1],syllables[1:])
        freq_bigrams_all = Counter(bigrams_all)

        tp_bigrams_all = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_all.items())

        print "Syllabes: %d, Bigrams: %d, Unique bigrams: %d" %(len(syllables), len(bigrams_all),
len(freq_bigrams_all)),

        bigrams_int = [bigram for word in words for bigram in zip(word[0:-1],word[1:])]

        freq_bigrams_int = Counter(bigrams_int)
        tp_bigrams_int = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_int.items())

        bigrams_acr = [(word1[-1], word2[0]) for (word1, word2) in zip(words[0:-1], words[1:])]
        freq_bigrams_acr = Counter(bigrams_acr)
        #freq_bigrams_acr = freq_bigrams_all - freq_bigrams_int
        tp_bigrams_acr = dict((bigram,float(freq)/freq_syll[bigram[0]]) for bigram,freq in
freq_bigrams_acr.items())
```

```
        aveacr = sum(tp_bigrams_acr.values())/len(tp_bigrams_acr) if len(tp_bigrams_acr)!=0 else
0

        word2 = dict()
        for (a,b) in words2:
                word2[(a,b)] = tp_bigrams_all[(a,b)]
        aveword2 = sum(word2.values())/len(word2) if len(word2)!=0 else 0
        len2 = len(word2)

        fc_words2 = Counter(bigrams_all)
        c_words2 = fc_words2.keys()
        for t in word2.keys():
                del fc_words2[t]

        words2_equal_sylb = {}
        for w in c_words2:
                words2_equal_sylb[w[-1]] = [ ww for ww in c_words2 if w[-1] == ww[0] ]

        ph_all2 = 0
        ph_frequent2 = 0
        for (a,b) in fc_words2:
                v = tp_bigrams_all[(a,b)]
                if v > aveacr:
                        ph_all2+=1
                if v > aveword2:
                        ph_frequent2+=1

        #print "Word2: ave = %.6f\tlen = %d\tph_all = %-12d\t ph_frequent = %d" %(aveword2,
len2, ph_all2, ph_frequent2)
        print aveword2,
        print len2,
        print ph_all2,
        print ph_frequent2,"\t",


        word3 = dict()
        for (a,b,c) in words3:
                word3[(a,b,c)] = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)])/2
        aveword3 = sum(word3.values())/len(word3) if len(word3)!=0 else 0
        len3 = len(word3)

        def words_next(words_prev_f, eq_sylb_2, word_exc):
                words_n = Counter()
                for x in words_prev_f:
                        for y in eq_sylb_2[x[-1]]:
                                z = x + (y[-1],)
                                if z in word_exc:
                                        continue
                                words_n[z] += 1
                return words_n.keys(), words_n


        c_words3, fc_words3 = words_next(c_words2, words2_equal_sylb, word3)

        ph_all3 = 0
        ph_frequent3 = 0
        for (a,b,c) in fc_words3:
```

```
		v = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)])/2
		if v > aveacr:
			ph_all3+=1
		if v > aveword3:
			ph_frequent3+=1

	#print "Word3: ave = %.6f\tlen = %d\tph_all = %-12d\t ph_frequent = %d" %(aveword3,
len3, ph_all3, ph_frequent3)
	print aveword3,
	print len3,
	print ph_all3,
	print ph_frequent3, "\t",


	word4 = dict()
	for (a,b,c,d) in words4:
		word4[(a,b,c,d)] = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)] +
tp_bigrams_all[(c,d)])/3
	aveword4 = sum(word4.values())/len(word4) if len(word4)!=0 else 0
	len4 = len(word4)

	#fc_words4 = Counter([(a,b,c,e) for (a,b,c) in c_words3 for (d,e) in c_words2 if c == d ])
	#c_words4 = fc_words4.keys()
	#for t in word4.keys():
	#	del fc_words4[t]

	c_words3.extend(word3)
	c_words4, fc_words4 = words_next(c_words3, words2_equal_sylb, word4)

	ph_all4 = 0
	ph_frequent4 = 0
	for (a,b,c,d) in fc_words4:
		v = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)] + tp_bigrams_all[(c,d)])/3
		if v > aveacr:
			ph_all4+=1
		if v > aveword4:
			ph_frequent4+=1

	#print "Word4: ave = %.6f\tlen = %d\tph_all = %-12d\t ph_frequent = %d" %(aveword4,
len4, ph_all4, ph_frequent4)
	print aveword4,
	print len4,
	print ph_all4,
	print ph_frequent4, "\t",


	"""
	word5 = dict()
	for (a,b,c,d,e) in words5:
		word5[(a,b,c,d,e)] = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)] +
tp_bigrams_all[(c,d)] + tp_bigrams_all[(d,e)])/4
	aveword5 = sum(word5.values())/len(word5) if len(word5)!=0 else 0
	len5 = len(word5)

	c_words4.extend(word4)

	ph_all5 = 0
	ph_frequent5 = 0
```

```
	for w in c_words4:
		c_words5, fc_words5 = words_next([w], words2_equal_sylb, word5)
		for (a,b,c,d, e) in fc_words5:
			v = (tp_bigrams_all[(a,b)] + tp_bigrams_all[(b,c)] + tp_bigrams_all[(c,d)] +
tp_bigrams_all[(d,e)])/4
			if v > aveacr:
				ph_all5+=1
			if v > aveword5:
				ph_frequent5+=1

	#print "Word5: ave = %.6f\tlen = %d\tph_all = %-12d\t ph_frequent = %d" %(aveword5,
len5, ph_all5, ph_frequent5)
	print aveword5,
	print len5,
	print ph_all5,
	print ph_frequent5, "\t"
	"""


print "Processing time: %.2f secs" %(time.time() - start)
print "Done."
sys.exit()
```