



PERGAMON

Journal of Neurolinguistics 17 (2004) 121–145

www.elsevier.com/locate/jneuroling

**Journal of
NEUROLINGUISTICS**

Indeterminacy in language acquisition: the role of child directed speech and joint attention

Peter F. Dominey*, Christelle Dodane

Institut des Sciences Cognitives, CNRS UMR 5015, 67, Boulevard Pinel, 69675 Bron Cedex, France

Received 28 February 2003; received in revised form 28 May 2003; accepted 24 June 2003

Abstract

Language acquisition represents one of the great learning achievements in human cognitive development. Perhaps, this process takes place in a relatively automatic manner in which, simply through exposure to language input, the child configures her language organ to coincide with the structure of the maternal language. In this context, the problem of the vast uncertainty between speech input and its external referent, related to the more general notion of the ‘poverty of the stimulus’ problem, takes on a significant importance, and motivates the nativist suggestion that language is already essentially preprogrammed, and acquisition consists of setting the parameters for the target language based on limited exposure.

What if, however, the acquisition process was not so automatic, but rather was controlled by the operation of mechanisms that could direct the attention of the child to specific aspects of the sentence and its external referent? In this case, external and internal control of attention could significantly reduce the referential uncertainty, thus reducing the requirement for preprogrammed language.

The current paper outlines evidence for this second scenario, in which child directed speech guides the child’s attention to important aspects of the speech signal, and Joint Attention focuses his attention on the relevant aspects of the referential world, significantly reducing the poverty of the stimulus problem. Results from recent simulation studies are briefly reviewed that indicate how these mechanisms could then allow a relatively non-specific learning mechanism to acquire initial knowledge of grammatical constructions in the first steps of language acquisition.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Language acquisition; Social interaction; Native language

* Corresponding author. Tel.: +33-437-911212; fax: +33-437-911210.

E-mail address: dominey@isc.cnrs.fr (P.F. Dominey).

1. Introduction

Language acquisition represents, perhaps, the most impressive achievement in human development. This is all the more fascinating since this process is quite rapid, and the successive stages for the progressive acquisition of the native language follow a quite similar chronology across languages. This achievement becomes increasingly impressive when considered from the perspective of the ‘poverty of the stimulus’ argument. This position holds that the linguistic input to the child, and its mapping onto meaning, is so under-specified (or poor in information), that language could never be ‘learned’ (nor a grammar induced) from such impoverished input. Indeed, spontaneous speech includes only a finite and limited number of sentences whereas the recursive property of language allows an unbounded generative capacity (Hauser, Chomsky, & Fitch, 2002). Moreover, spoken language is degraded by interruption, hesitation, repetition and the use of non-grammatical utterances, and few or no meta-linguistic indications are available to help the child to determine which sentences are non-grammatical. A related problem is that when an infant hears an utterance, and perceives a visual scene, it is not at all obvious to what aspect of the scene the sentence is referring, nor the perspective taken by the speaker (Quine, 1960). Thus, the problem of mapping sentence to meaning is under determined, or indeterminate, and ‘learning’ language appears to be impossible.

Nevertheless, and despite the supposed deficient input, normal children succeed in learning to speak, in a very effective manner. According to nativist theory (Chomsky, 1965, 1995), this achievement is explained by the presence of a powerful innate language faculty. In this context, the child is endowed with a universal grammar that becomes adapted to the particular target language through exposure. The problem of the poverty of the stimulus is solved by the fact that the language organ or language acquisition device is already in place and that very little learning is required. Still, language acquisition will require that the infant can represent the input signal, and the structural information contained in the input signal (Chomsky, 1965).

From this perspective then, one might ask, what is the structural information, and to what extent is the language acquisition environment so impoverished. Indeed, according to Morgan (1986), the input is not impoverished, but on the contrary, carries important cues on syntactic grouping that reduce the requirements on the complexity of the underlying innate grammar. In this context, the prosodic bootstrapping hypothesis (Gleitman, Gleitman, Landau, & Wanner, 1988; Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Morgan, 1986; Morgan & Demouth, 1996; Morgan, Meier, & Newport, 1987; Morgan & Newport, 1981; Peters, 1983) proposes that the prosodic structure of the speech signal contains rich information concerning the syntactic structure of the language that can aid the infant in initiating the learning of this structure. In this developmental perspective, the acquisition of a small lexicon of concrete nouns independent of grammatical knowledge (Brent & Siskind, 2001; Siskind, 1996) can provide a further basis from which syntactic categories and structures can then begin to be derived by the pairing of actions in situational contexts (meanings) with syntax (forms) through semantic bootstrapping (Gillette, Gleitman, Gleitman, & Lederer, 1999; Pinker, 1984, 1987, 1994). In turn, the growing syntactic knowledge can then be used to make inferences about word meanings via syntactic bootstrapping (Gleitman, 1990).

The interaction between these mechanisms will be considered in more detail below. Further extending the emphasis on learning in language acquisition, a growing number of theoretical and experimental studies are developing a position in which social interaction and learning plays a much more central role in language acquisition (Carpenter & Tomasello, 2000; Goldberg, 1995; Kuhl, 2000; Langacker, 1991; Tomasello, 1999, 2003).

However, even if learning is to assist or partially replace the capabilities of the highly specialized, pre-wired language organ, there still must be a solution to the indeterminacy and poverty of the stimulus problem. Again, the indeterminacy is that it is not known for a given ⟨sentence, meaning⟩ pair, which words and grammatical structures correspond to which aspects of the meaning. Indeed, even negotiating agreement on the meaning shared by speaker and listener is a vast problem. The possible mappings are potentially infinite. Thus, the ability to focus the attention of the infant on a particular word, and on a particular object in the scene could significantly contribute to the resolution of this problem. Indeed, this indeterminacy in the sentence to meaning mapping can be reduced in at least two distinct and powerful manners.

First, indeterminacy can be reduced in the sentence component by directing the child's attention to the appropriate aspect of the elements of the sentence through modulation of the acoustic signal, as is typically the case in child directed speech (CDS). Nativist theory tends to minimize the importance of this type of language (Crain & Lillo Martin, 1999). We will review studies indicating that there is significant information in the CDS signal, and that even without CDS, there is still significant extra-lexical linguistic information in the acoustic speech signal that aids in reducing the poverty of the stimulus. Second, indeterminacy can be reduced in the meaning component by focusing or directing the child's attention to the appropriate aspect of the visual scene through the creation of joint attention interactions between the child and the caregiver.

In the rest of the paper we will review data that indicate that indeed, the attention of the child is significantly engaged by both of these mechanisms, thus rendering language more learnable. The review will not be exhaustive, but rather attempts to identify a subset of illustrative examples. Part of what we will consider is that the attentional cues should fulfill four criteria. The cues should be present in the signal, infants should be sensitive to the cues, use of the cues should lead to improved performance, and failure to use the cues should lead to negative influence on performance.

2. Modulation of attention in child directed speech

Often, when the poverty of the stimulus argument is being developed, the sentence is considered formally as a sequence of words, and little attention is paid to the possible information content of the acoustic structure of the speech signal (Crain & Lillo-Martin, 1999). In reality, contrary to this perspective, the acoustic structure the speech signal is not poor, but rather contains significant information concerning the structural organization of the language (Morgan & Demouth, 1996; Morgan et al., 1987; Morgan, Shi, & Allopenna, 1996). To the extent that this source of information is rich, the poverty of the stimulus argument is weakened, and the infant can exploit this information to infer the appropriate corresponding syntactic properties of her maternal language.

According to the prosodic bootstrapping or ‘signal bootstrapping’ hypotheses (Gleitman et al., 1988; Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Morgan, 1986; Morgan & Demouth, 1996; Morgan et al., 1987; Morgan & Newport, 1981; Peters, 1983), prosodic structure indicates different levels of linguistic structural organization (e.g. sentences, clauses, words). By exploiting this information, the child is significantly aided in inducing knowledge of the syntactic organization of her native language. Indeed, Read and Schreiber (1982: 26) note that “an infant who is innately biased to treat intonationally circumscribed utterance segments as potential syntactic constituents, would be at considerable advantage in learning the syntactic rules of his language”. This hypothesis assumes that the signals should be sufficiently perceptually salient to be perceived by the infant. Effectively, as reviewed below, a number of studies indicate that prosodic cues are particularly salient in the speech signal and correspond to different level linguistic unit boundaries (Morgan et al., 1987), from higher level (phrases, sentence) to more lower level (words).

In English, phrasal-sentence frontiers are marked by a pause and an increasing/decreasing pitch contour in addition to the lengthening of the clause-final syllable (Klatt & Cooper, 1975; Luce & Charles-Luce, 1983). Likewise, syllables and segments which end clauses and sentences tend to be lengthened (Cooper & Paccia-Cooper, 1980). Moreover, this lengthening is exaggerated in CDS (Bernstein Ratner, 1986; Morgan, 1986). Pauses and F0 modulation represent strong cues to facilitate the localization of clause boundaries. Moreover, these cues are strongly present in CDS (Fernald & Simon, 1984) and, for example, in stories read to young children (Morgan, 1986). The infant seems to be particularly sensitive to these clausal cues (Hirsh-Pasek et al., 1987; Jusczyk, 1989). In a study addressing infant sensitivity to the predicate–subject difference (Section 2.4), Jusczyk et al. (1992) revealed that pitch changes and durational changes are potential cues for the localization of the predicate–subject boundaries.

Gleitman and Wanner (1982) assume that stressed syllables could play a major role in bootstrapping learning. Because of their perceptive salience, stressed syllables are probably easier to localize. This information in syllable stress could help the child in lexical categorization. For example in English, function and content words can be differentiated with stress (Gleitman & Wanner, 1982; Morgan et al., 1996) and disyllabic nouns tend to receive stress on initial syllables, whereas disyllabic verbs tend to receive stress on final syllables (Kelly, 1988). In fact, for English children, the learning of non-stressed syllables is slower than the learning of stressed syllables (Brown, 1973; Gerken, 1996). Thus, according to Gleitman and Wanner (1982), stress could be a good candidate to distinguish function and content words. According to Selkirk (1996), if infants are able to make a strong/weak difference, they could discover relations between these differences and word categories. This issue has been explored by Morgan et al. (1996) who demonstrated indeed that distributional, phonetic and acoustic cues taken together provide a rather robust lexical categorization capability. In an investigation of how such information in ADS could be used in early language acquisition, we examined how properties of the fundamental frequency (F0) profile are correlated with lexical categories in French and English. In particular, we observed that the presence of local peaks in the F0 signal predicted that the word containing the peak was an open class (content) word, while absence of an F0 peak predicted that the word was a closed class (function) word (Blanc, Dodane, & Dominey, 2004).

Thus, multidimensional information concerning different levels of structural organization (e.g. sentences, phrases, lexical categories) is carried by a constellation of cues in the speech signal, and can thus contribute to the orientation of the child's attention to different levels of structural organization of the native language.

2.1. *Child directed speech*

Interestingly, these prosodic signals are specifically exaggerated when an adult addresses a child (Fernald & Mazzie, 1991; Jusczyk et al., 1992; Lederer & Kelly, 1991; Morgan, 1986). From birth, the mother speaks to her child in a particular manner, with simplified syntax and meaning, and exaggerated prosodic structure as if adapted to the capacities of the child. In this CDS, 'motherese' (Ferguson, 1964) or 'parentese', the segmental structure is deformed in favor of the exaggerated prosodic structure, yielding the highly musical properties of this 'sing-song' language form (Fernald, 1989; Papousek & Papousek, 1981). In this way, the information conveyed by prosodic cues is emphasized and adapted to the infant's perceptual capabilities. These vocal adjustments are made in an intuitive and natural manner, and are rarely used outside the context of interaction with a child (Papousek & Papousek, 1987). Thus, in this context caregivers appeared to provide 'a real tutorial and they did not even know they were doing it' (Fernald, 1976).

2.2. *Exaggerated acoustic marking of constituent structure*

While the majority of CDS sentences are short (Broen, 1972; Newport, Gleitman, & Gleitman, 1977) and often grammatically simplified or incomplete, they tend to be complete from the perspective of prosodic phrase structure (Broen, 1972). Likewise, mothers tend to insert more pauses between sentences and propositions, and these pauses are lengthened with respect to those in ADS (Broen, 1972; Fernald & Simon, 1984). Thus in English and Japanese, syllables preceding prosodic boundaries are acoustically distinctive, with long pauses preceded by elongated vowels, accentuated by exaggerated F0 contours (Fisher & Tokura, 1996). Vowels preceding the pauses are twice as long, with an F0 profile twice as high as those vowels in other sentence positions. Likewise, Kuhl et al. (1997) observed that American, Russian and Swedish mothers when speaking to their 2–5 month old babies produce more lengthened vowels and use prosody to attract the babies attention to the three vowels [i], [a] and [u]. Moreover, these vowels are acoustically more extreme in CDS than in ADS.

These characteristics of CDS are not static, and the mother adapts her speech to the developing child. During the four first months, mothers insert more and more isolated words, pauses, and breaks (Broen, 1972). Then, when the baby reaches approximately 4 months, these characteristics are attenuated. From four through six months, the infant–mother interaction is the most 'intensive'. The mother uses rising pitch contours to attract her baby's attention and bell shaped pitch contours to maintain attention and eye-contact with her (Stern, Spiker, & McKain, 1982). Later, when the baby reaches ~14 months, the mother stresses the predicate by employing a higher pitch and a greater amplitude. In parallel, new words are localized in final position (Aslin, Woodward, LaMendola, & Bever, 1996) and are lengthened and pronounced with a higher pitch (Fernald & Mazzie,

1991). At the same time, the baby's comprehension abilities for words in her native language are rapidly increasing. By the age of 8 months, the baby is able to recognize words out of context, and her attention focus longer on familiar words, even if they are phonologically altered (Jusczyk & Aslin, 1995), and at 12-months, the first isolated words appear within mixed babbling (De Boysson-Bardies & Vihman, 1991).

According to Kuhl et al. (1997), vowels and consonants are pronounced more distinctively when she child begins to speak (hyperarticulation) than when she is younger or older (Malsheen, 1980). Motherese is a way to adapt speech to the child's perceptive abilities and thus must evolve as a function of her development: 'to be effective, the communicative signals of the mother must be well matched to the perceptual capabilities and limitations of the young' (Fernald, 1989: 10). This creates a form of 'mirror game' in which the mother reflects the processing capabilities of the child, and emphasizes the important linguistic units in the speech signal and continues to do so until she perceives that the baby is ready to pass on to a more difficult step in acquisition of her language. This accentuation of syntactic structural units in CDS by prosodic exaggeration provides an enhanced learning context for the infant. Identification and subsequent exploitation of these perceptually salient units should thus be facilitated.

2.3. Use of intonation and rhythm to modulate interaction with the infant

Another function of prosody in CDS is to regulate the child–caretakers' interaction, and F0 contours seems to play an important role in this function. In fact, the most salient and specific characteristic of CDS is the raising of the voice pitch by 3–4 half tones (Fernald & Simon, 1984; Fernald et al., 1989; Papousek & Papousek, 1981; Stern, Spieker, Barnett, & McKain, 1983). After isolating the acoustic parameters differentiating CDS and adult directed speech (ADS), Fernald and Simon (1984) observed a significant continuity in the temporal evolution of the pitch (melody) with respect to its characteristic variability observed in ADS. The contours are quite exaggerated, but their configuration remains simple and uni-directional (high–low, or low–high), rendering them easier to process for the child. "The exaggerated intonation of mothers' speech, with its greatly expanded pitch range and high degree of pitch continuity, thus maximizes both perceptual contrast, necessary for engaging and maintaining infant attention, and perceptual coherence, facilitating the task of following the voice of a single speaker" (Fernald, 1984: 13). Papousek (1995) developed a typology of nine prototypical contours that recur regularly in CDS. The contours are adapted to the attentional state of the baby, with exaggerated and elevated pitch ranges when the infant is alert, descending contours to signal interdiction, and rising contours to solicit attention and to incite the child to participate in the dialog. Inversely, the more the attentional state is reduced, the more the profile is soft and descending to calm the child. In this way, F0 contours permit modulation of awakening and soothing functions.

2.4. Infant sensitivity to prosodic cues

This brief review indicates that certain acoustic–syntactic correlations are available in the signal, but a prosodic bootstrapping theory requires both that the speech signal

must contain this syntax-related information, and that the infant must be capable of perceiving and exploiting these cues. Indeed, a number of psycholinguistic studies have demonstrated the sensitivity of infants to these cues. Thus, concerning voice pitch, babies are capable of discriminating identical syllables that vary only in their rising and falling intonational contours (Kuhl & Miller, 1982; Morse, 1972). Menyuk (1994) obtained similar results with infants at 6–7 weeks of age. Already at 4 days after birth, infants are capable of identifying prosodic boundaries that differentiate the bisyllabic group ‘mati’ in ‘panorama typique’ et ‘mathématicien’ (Christophe, Dupoux, Bertocini, & Mehler, 1994). Interestingly, Best (1993) demonstrated that English babies distinguished prosodic contrasts in English, but only for CDS. In parallel, these infants are able to respond appropriately to the underlying message such as a request to wake up, to pay attention and to look (Papousek, Bornstein, Nuzzo, Papousek, & Symmes, 1990).

With respect to speech rhythm, infants can discriminate between sequences with different rhythmic structures from 2 to 3 months of age (Chang & Trehub, 1977; Demany, McKenzie, & Vurpillot, 1977; Morrongiello, 1984). Newborns and infants can discriminate between multisyllabic stimuli that differ only in the location of the accent (Jusczyk & Thompson, 1978; Sansavini, Bertocini, & Giovanelli, 1997; Spring & Dale, 1977). In this same period, infants are sensitive to heavily accented words and sentence endings (Menyuk, 1977). In contrast, they are insensitive to words pronounced without prosodic exaggeration (Eimas, 1975; Kagan & Lewis, 1965). Using low pass filtering, to retain primarily the rhythmic structure, Jusczyk et al. (1993) demonstrated that 6-month-old English babies could discriminate English words from Dutch. This observation was subsequently extended to show that newborns can discriminate between foreign languages from different rhythmic classes, again based on filtered speech that preserves the rhythmic structure (Nazzi, Bertocini, & Mehler, 1998).

Not only are infants sensitive to these prosodic regularities, they appear to be able to use them in segmentation from the age of 4 months. They listen longer to speech with pauses inserted at phrase boundaries than to those with pauses inserted within phrases (Jusczyk, 1989). Hirsh-Pasek et al. (1987) likewise demonstrated that babies at 6 months prefer CDS extracts when pauses are inserted at natural boundaries. It should be stressed that this preference is displayed only with CDS (Kemler Nelson et al., 1989), indicating that CDS is particularly attractive for the child.

Clearly there exist other cues in continuous speech to which infants are sensitive, such as statistical or distributional regularities, though prosody appears to predominate. For example, Saffran, Aslin, and Newport (1996) showed that 8-month-old infants are sensitive to statistical segmentation regularities in a continuous sound stream. Mattys, Jusczyk, Luce, and Morgan (1999) subsequently demonstrated that when statistical and prosodic cues are used in combination, the prosodic cues take precedence as infants prefer the predominant English strong/weak stress pattern in stimuli that violated phonotactic cohesion. Similarly, Johnson and Jusczyk (2001) demonstrated that speech cues (stress and coarticulation) carry more weight than statistical cues for 8-month-old infants in a Saffran-like task.

2.5. Preference for CDS

If the child appears to be sensitive to prosodic cues, he should prefer CDS because these cues are exaggerated in comparison to ADS. Indeed, this preference for CDS appears well before birth, as it has been displayed in the fetus (Nowik-Stern, Clarkson, Morris, & Bakeman, 1998). Babies equally prefer CDS (Cooper & Aslin, 1990; Fernald, 1985; Werker & McLeod, 1989) even when the contours are synthesized with sine waves (Fernald & Kuhl, 1987), and infants indeed display signs of pleasure when exposed to CDS (Papousek & Papousek, 1981). Fernald and Kuhl (1981) were able to isolate the acoustic parameters that distinguish CDS from ADS. The preference for CDS appears to be based uniquely on the modifications of F0, a parameter that is likely more salient than the rhythm, amplitude and higher formants. Children thus preferred speech in which the F0 varied (Fernald, 1985; Fernald & Kuhl, 1987) with rapidly rising contours (Fernald, 1989).

2.6. Influence of CDS on learning

The child's sensitivity to prosodic cues and her preference for CDS over ADS highlight the potential importance of CDS during the acquisition process. If CDS plays an important role in guiding the child's attention to relevant aspects of speech, what happens when these prosodic characteristics are altered? When depressive mothers address their children, their speech displays an F0 that is less modulated than in typical CDS (Kaplan, Bachorowski, Smoski, & Zinser, 2001). Kaplan, Bachorowski, Smoski, and Hudenko (2002) observed that this loss of prosodic emphasis rendered the depressed CDS less effective in controlled learning tasks. However, the children were able to learn when exposed to normal CDS. These results suggest that the loss of perceptual salience in CDS could have a negative impact on learning.

If an impoverished CDS can disturb the acquisition, it does not necessarily imply that a child deprived of CDS could not acquire his native language. The question is to know if CDS is necessary for the acquisition of language. According to Monnot (1999), CDS is a universal feature common to modern humans. CDS contributes to the acquisition of language as well as emotional regulation and socialization. But even if CDS is observed in essentially all languages of the world, there exist particular known exceptions including Samoan and Quichee Mayan in Guatemala (Pye, 1986). This apparent weakening of the CDS argument, i.e. that, non-CDS languages are still learnable, actually raises a quite interesting point. The essential acoustic property of CDS is the exaggeration or modulation of characteristics that are *already present* in ADS. This suggests that the functional characteristic of CDS, to aid in the identification of significant aspects of the signal, is also characteristic of ADS, though in a reduced form. That is, in a certain sense, all language is 'child-exploitable' because of the multidimensional structured extra-lexical information that is available in the acoustic signal to provide cues about the structural organization of the language.

The infant is thus in a privileged learning environment in which different modalities converge to facilitate learning. In exaggerating the relevant information in CDS, caretakers increase the accessibility of these cues that are already present in adult speech. The simultaneous use of other modalities including gestures, touch, facial

expressions and eye-contact, interact to make available information highly redundant. In particular, joint attention allows the focusing of the child's attention on relevant components from the referential word. From Falk's (2003) perspective, mothers establish eye-contact unconsciously with their child and maintain their attention via CDS. This suggests that language acquisition could be considered as a learning process taking place in an extremely rich interactive context.

3. Joint attention and language acquisition

In CDS, caregivers modify the normal prosodic structure of speech in order to draw the infant's attention to relevant aspects of the speech signal. Similarly, the caregiver can also draw the infant's attention to relevant aspects of the surrounding world so that both are jointly attending to a common aspect of the situation. It is clear that such a mechanism of joint attention will play a powerful role in reducing the indeterminacy of the intended referent or meaning during exchanges between the caregiver and the child. Bruner (1975) proposed that interaction between speaker and child could enable the setting up of pre-linguistic meanings that would satisfy or ground the referential uncertainty problems exposed by Quine (1960). In this context, joint attention thus refers to the ability of an infant to coordinate his/her attention with a social partner with respect to an object or event, and has a long history of inquiry in child development (Bruner, 1975; Moore & Dunham, 1995; Scaife & Bruner, 1975; Tomasello, 1988). Engaging in joint attention is thought to contribute to the process of language acquisition because these types of object-mediated social interactions help infants identify the intended referent of parents' language, thus facilitating word-object mappings.

3.1. Presence of joint attention cues

The solicitation of joint attention is an integral component of human social interaction. Joint attention interactions between adult caretakers and infants often take place during routine daily activities such as bathing, feeding and book reading that provide additional contextual regularity (Moore & Dunham, 1995; Tomasello, 1992). Joint attention can be solicited in different ways that include pointing to or physically manipulating the object in question, and directing gaze to the object or event. In a series of ecological studies that involved recording and analysis of joint attention interactions between children and caregivers, Tomasello and colleagues (Carpenter, Nagel, & Tomasello, 1998; Tomasello & Todd, 1983) observed the natural interaction between mothers and their infants. There they noted that mothers could use joint attention in two distinct manners. They could follow their child by talking about an object that was already at the focus of the child's attention, or they could direct the child, using language to change the child's focus of attention to a different object or event. Interestingly, children whose mother's had an increased tendency to follow their attention had larger vocabularies than those whose mother's tended to direct their attention. These and related studies support the strong intuition that joint attention cues are present, and indeed play an important role in

interpersonal human interaction (see [Moore & Dunham, 1995](#) for a more extensive treatment of these issues).

3.2. *Sensitivity to joint attention cues*

In this context, a crucial aspect of the development of the child's manner of relating to the social world during the first year of life is the transition from engaging exclusively in dyadic interactions (either with people or objects) to taking part in triadic exchanges at around their first year ([Tomasello, 1992](#)). In these exchanges, the triad is made up of the infant and the caregiver who are in a joint attention context, and a third object or event that is the focus of their attention. In this context, the capacity to follow the direction of gaze, or the ability to respond to the joint attention (RJA) bids of others, is one of the earliest indications of this capability ([Scaife & Bruner, 1975](#)). There is considerable debate however, concerning the developmental period at which this capability appears.

A number of studies have contributed to this debate. With respect to the age at which this process can begin to be observed, [Morales et al. \(2000b\)](#) observed that indeed, under the appropriate conditions, infants as young as 6 months can display RJA behavior. They consider that as the 'ecological validity' of the paradigm used to evaluate RJA is reduced, so is the probability of finding RJA behavior before 10 months of age. In particular they considered the following three factors as contributing to the ecological validity of the paradigm: Whether or not the JA partner is a familiar parent, whether vocal as well as visual gaze cues are used, and finally, whether or not the infants have been made aware of the existence of the RJA targets before the experiment begins. This third point is of particular interest when compared with the results on the ability of infants to imitate an observed behavior. [Carpenter, Call, and Tomasello \(2002\)](#) observed that infants extracted more information from observation of actions in an imitation task when they had some prior knowledge of the intention of the person they were observing.

Interestingly, it appears that these attentional mechanisms can also be applied to the interpretation of emotional messages. That is, the importance and relevance of emotional cues (e.g. surprise, pleasure) are determined based on the interpretation of referential, attentional cues. Thus, infants at 12 and 18 months reliably use referential attentional cues including gaze and postural orientation to evaluate whether an emotional message should be linked or not with a salient novel object. These children will also actively exploit referential cues in order to disambiguate the intended target of an emotional display ([Moses, Baldwin, Rosicky, & Tidball, 2001](#)). Such results indicate that infants are sensitive to the joint attention cues in their environment.

3.3. *Responsiveness to joint attention bids improves language acquisition*

In the proposed argument that joint attention will reduce referential indeterminacy, a crucial point will be to demonstrate a functional link between joint attention capability and future language capabilities during development. Indeed, a number of studies have directly addressed this point, and the evidence appears to be in favor of a positive correlation between early RJA capabilities and later language skills. For example, [Morales et al. \(2000a\)](#) observed that significant associations were found between 6-, 8-, 10-, 12-, and

18-month RJA measures and vocabulary acquisition. In particular, RJA at 6, 8, and 10 months was significantly and positively correlated with receptive vocabulary at 30 months. Likewise, there were a number of significant associations between RJA and expressive vocabulary at 24 and 30 months (Morales et al., 2000a). Likewise, longitudinal associations between joint attention abilities (including proto-declarative pointing, following gaze, and following pointing) and later language ability in the second year of life have been reported by Carpenter et al. (1998). In this same line of research it has been demonstrated that mother–infant dyads who spent more time in JA engagement had infants who used more gestures and understood more language in the following months (Carpenter et al., 1998).

In related studies, joint attention parameters were manipulated in order to determine the effects on word learning. Thus, Dunham, Dunham, and Curwin (1993) tested the ability of 18-month old children to learn new words in an attention following context in which the words were presented when the child was attending to the referent, and in an attention switching context, in which words were presented while the infant was attending elsewhere and had to switch attention. This study revealed a significant learning advantage for the attention following vs. attention switching context. In a related study, Harris, Jones, Brookes, and Grant (1986) demonstrated that the mothers of children with slowed language development initiated more changes in conversation topic without providing appropriate non-verbal attentional clues; they made more references to objects outside the attentional focus, and less to those with the attentional focus of their children; and they used fewer specific labels, and more general terms (e.g. pronouns) than did mothers of children with normal language development. These data suggest that the ability to exploit joint attention cues may be a useful predictor for future language development, and that the reduction of these cues can lead to slowed language development.

3.4. Failure to respond to joint attention is correlated with language dysfunction

The final criteria for establishing the functional link between joint attention capabilities and language capabilities is the demonstration that language learning is impaired in children that display dysfunctions in joint attention processing. In this context, the study of language behavior in childhood autism has played a crucial role. Autism is a developmental disorder that is characterized by a restricted scope of activities and interests, and impairments in social interaction and communication capabilities. Impairments in joint attention are specific et almost universal in children with autism, and these effects have been extensively well replicated (Dawson et al., 2002; Mundy, 1995). These impairments can be observed at 1 year of age (Mundy, Sigman, Ungerer, & Sherman, 1986; Osterling & Dawson, 1994), and it appears that these deficits are correlated with subsequent language development dysfunctions. Thus, when controlling for language level, mental age, and IQ, autistic children displayed deficits in gestural joint attention skills on two testing sessions that were 13 months apart. Most importantly, the measure of gestural non-verbal joint attention was a significant predictor of language development in the autistic sample (Mundy, Sigman, & Kasari, 1990).

Indeed, given the important role that joint attention is claimed to play in language acquisition, we should expect that autistic children should display clear relations

between the joint attention and language dysfunctions. Thus as described above, correlations between measures of joint attention and measures of children's language are consistently found (Carpenter & Tomasello, 2000; Mundy et al., 1990). More generally, a deficit in communicative use of language is among the defining characteristics of autism in the DSM IV. The deficits in these children on the autistic spectrum are similar to those of children with developmental language disorders, though the autistic children differ markedly by their universally impaired pragmatics and discourse comprehension. From a functional point of view, Baron-Cohen, Baldwin, & Crowson (1997) demonstrated that autistic children fail to exploit the speaker's direction of gaze direction as an index of the speaker's intention to refer. The data from these studies of autism thus support the proposition that JA plays a central role in language acquisition by engaging the child in a triadic relation around an event or object, thus reducing the problem of referential uncertainty. In the absence of this JA capability in autism, language acquisition is impaired.

The question of the necessity of visual joint attention remains open, however, including the relative contributions of different types of joint attention. The term joint attention can refer to both the child's response to joint attention 'bids' (RJA) that have been initiated by the child's caregiver, as well as to the initiation of joint attention behavior by the child herself (IJA). What is the relative contribution of these types of behavior? The short answer is that both types of these types of joint attention behavior contribute to the formation of the triadic behavior of the child and caregiver, organized around the external referent in question. To the extent that any behavior contributes to the formation of these attentional triads, it will reduce referential uncertainty and facilitate language acquisition. Now one might ask, what happens in cultures that do not engage in much JA with infants: they learn to speak, but how? How can referential uncertainty be reduced? Baldwin (1991) demonstrated that already at 16–19 months, infants can follow the gaze of adult speakers to resolve referential uncertainty, even when that gaze is directed away from where the child is looking, and thus does not directly engage the child. This suggests that reading such attentional cues allows the child to determine the likely target of what adults are 'talking about.'

Another factor that can aid the child to infer what someone is talking about is the child's knowledge of the speaker's intentions. In this context, Baldwin, Baird, Saylor, and Clark (2001) have experimental evidence that 11-month-old infants are sensitive to disruptions in the spatiotemporal structure inherent in intentional action, and thus that they parse ongoing behavior with respect to this structure. The authors argue that this parsing skill is likely a prerequisite to the development of genuine intentional understanding. Understanding the speaker's intentions will clearly help to reduce the referential uncertainty with respect to what the speaker is talking about. Indeed, though it is beyond the scope of the current article, we begin to see the developmental continuum from the ability to exploit direct joint attention cues, via the ability to infer agent's sensorimotor intentions, leading finally to the development of a full blown theory of mind in which the child can attribute intentions, goals and mental states to the speaker (Leslie, 1994)—all of which will contribute to the reduction of referential uncertainty.

4. Theoretical implications

Experimental manipulation of CDS and JA in children is technically difficult for a number of reasons. Ideally, these effects and related aspects of language acquisition could be studied in computational models in which various parameters could be freely manipulated. In this context, the effects of speech modifications in CDS on the perceptual acquisition of vowels has been studied in a computational model of vowel learning. In CDS the acoustic formant patterns of centers of vowels are farther apart than those in adult-directed speech, which could help children to learn the categories of speech sounds of their language. In this context, [DeBoer and Kuhl \(2001\)](#) recorded words from mothers in both infant- and adult-directed speech, and extracted first two formants from the voiced parts of each word. Indeed, the infant-directed words used a larger part of the acoustic space, and also resulted in a better distribution of the centers of the vowel representations, indicating better patterning of the input, both of which make it easier to learn the categories of speech sounds.

Simulation studies have also been of use in determining the nature of the parameters of speech that contribute to early language acquisition. In this context, [Shi, Morgan, and Allopena \(1998\)](#) performed simulation studies to determine the relative contribution of different acoustic parameters to the ability to perform lexical categorization between open and closed class words. They determined that together, parameters including syllabic structure, duration and relative amplitude allowed an unsupervised learning system to perform this lexical categorization in Mandarin and Turkish at a level of greater than 80% classification accuracy performance. One limitation of these simulation studies is that the temporal acoustic parameters were recoded symbolically in order to be compatible with the input format of the model, thus introducing a significant potential loss of information. We have recently overcome this limitation, performing simulation studies in which the input to a ‘temporal recurrent network’ (TRN) was a continuous value of the F0 signal alone, sampled at 5 ms intervals, thus preserving the integrity of the temporal structure of the F0 contour. [Dominey and Ramus \(2000\)](#) had previously demonstrated that the TRN could exploit prosodic structure in consonant–vowel sequences in order to discriminate between languages from different rhythmic classes. The simulation results with the F0 temporal contour demonstrated that the TRN could perform lexical categorization of open and closed class words at levels of 70 and 62% classification performance for French and English speech, respectively ([Blanc et al., 2004](#)). Interestingly, these studies were performed using ADS. This indicates that, as suggested above, even ADS is ‘child-exploitable’ for lexical categorization. Future studies will examine if CDS renders this capability more robust.

Related simulation studies have also demonstrated the value of prosodic information in segmentation. [Shriberg, Stolcke, Hakkani-Tür, and Tür \(2000\)](#) examined the performance of automatic models for the segmentation of spoken language (speech from Broadcast news). Using decision tree and hidden Markov modeling techniques, they show the prosodic model (using timing and melody) works as well as or better than word-based statistical language models for the detection of sentence and topic boundaries.

Similarly, simulation studies have been used to test the effects of different attentional strategies on word learning. Such studies typically involve populations of agents involved

in different types of language games. A classical example is the talking heads study of [Steels and Kaplan \(2002\)](#) in which thousands of agents engaged in discrimination games involving two agents and form of joint visual attention that restricted their shared visual context. In the discrimination game, one agent would describe a physical object and another would attempt to determine to which object the speaker was referring. Over time, as successful descriptions were retained, agents acquired lexicons that allowed them to perform with progressively greater success. Addressing the issue of joint attention in a related experimental context, [Vogt and Coumans \(2003\)](#) studied three different attentional strategies (games) in a population of learners. In the observational strategy, meaning is transmitted between agents non-verbally via joint attention. In the guessing strategy, communication between agents is verified through the use of corrective feedback. In the third ‘selfish’ strategy, the hearer has no means of knowing the meaning of a word other than the co-occurrence of the word with a number of meanings that are present within the context of a game, corresponding to a non-joint attention context. Simulation experiments reveal that, though the selfish game can result in learning under certain conditions, there is a significant advantage to the joint attention based learning strategies. Such studies provide useful insight into the effects of these attentional mechanisms, and allow the simulation of different attentional strategies that is technically difficult with human subjects.

These studies provide support for the claim that JA and CDS can aid in learning words (at least vowels that will be useful in word discrimination), lexical categories and word meanings. In a larger theoretical perspective, we can consider that these mechanisms for reducing referential uncertainty will fall into a continuum of progressively more elaborated mechanism. In this context, it has been demonstrated that as children come to be familiar with the relations between word order in the simplest grammatical constructions, and the referential correspondence or meanings of the elements in these constructions, they can use this knowledge to reduce referential uncertainty in learning new words. More concretely, if a 2.5-year-old child is presented with a visual scene narrated by a sentence such as ‘The boy is gorging the ball,’ the infant will be able to use her acquired knowledge of the active transitive grammatical construction in order to infer that ‘gorging’ refers to the action that the boy is performing on the ball ([Hirsh-Pasek & Golinkoff, 1996](#)). That is, we are now passing into a domain in which structural knowledge of the language itself can be used to reduce referential uncertainty via syntactic bootstrapping, in which syntactic patterns in the input are used to make inferences about word meanings ([Gleitman, 1990](#)). In the reciprocal image, knowledge of individual word meaning can be used to establish the relation between grammatical structure and the meaning structure of the narrated scene, in a process of semantic bootstrapping ([Pinker, 1984, 1987, 1994](#)).

We have recently addressed these learning mechanisms in the context of a neural network model that has been used to describe and predict a number of psycholinguistic and neurolinguistic phenomena ([Dominey, Hoen, Lelekov, & Blanc, 2003](#)). Extending the model to more ecological meaning representations, we have performed simulation studies of acquisition of word meanings and grammatical constructions ([Dominey, 2000, 2003](#)). In these studies, sentence–meaning pairs are presented to the model, which attempts to learn word meaning and syntactic structure. The model is thus faced with indeterminacy at the word level, and the sentence level. Sentence input consists of successive words,

and meaning input consists of a conceptualized scene representation in the form (event(agent, object, recipient)), paired with the sentence. Based on these inputs, the system learns the mappings of individual words to their referents, and mappings of grammatical sentence structures to the meaning structures.

Like the child, the model is faced with the problem of how word meaning and grammatical structure can be learned, and crucially—the logical circle in which semantics relies on syntax, and syntax on semantics. An interesting approach is proposed by Gillette et al. (1999: 168) who state that “the circle is avoided because the novice breaks into the system (grounds the learning process) by a first-pass asyntactic analysis whose primary output is a concrete nominal vocabulary. This knowledge of nouns and then serves a second crucial purpose, for it underpins the construction of clause-level syntax—which in turn enables further vocabulary learning.”

Thus, in our simulations, word learning initially operates independently of grammatical learning corresponding to the asyntactic first pass of Gillette et al. (1999). Words are associated with all referents in the current scene meaning, or context, similar to the selfish strategy examined by Vogt. Based on the cross-situational regularity (Siskind, 1996) that a given word will co-occur more often with its referent than with any single non-referent, the system can learn a first approximation of the lexicon. The robustness of this strategy was explored and characterized in more elegant manner by Siskind (1996). This initial knowledge of word meaning can then be used to learn the mapping of sentence structure onto meaning structure at the scene or event level, corresponding to the construction of clause level syntax via semantic bootstrapping (Gillette et al., 1999). In our demonstration of this procedure, the ‘construction of clause-level syntax’ takes place via a sentence-to-meaning mapping mechanism that is distinctly non-UG-like in its character (Dominey, 2000, 2003). Rather, it is a structure mapping mechanism that builds Goldberg-like grammatical constructions (Goldberg, 1995) by establishing the mapping between sentence structure and event meaning structure. That is, the child can recognize that in some sentence types (e.g. the active transitive) the first noun (‘thing’) is the agent or doer, and the second one is the object. This knowledge in the form of a grammatical construction can then generalize in a fully productive manner to all sentences of this active transitive type, corresponding to the grammatical constructions described by Goldberg (1995). In the model, the mapping of grammatical constructions onto meaning requires that different grammatical constructions (e.g. active and passive forms of transitive and ditransitive constructions) can be uniquely identified, and associated with a specific form to meaning mapping. The identification of distinct grammatical constructions is based on specific constellations of closed class function words or morphemes that uniquely identify different grammatical constructions, as specified more generally in the cue competition model of Bates and MacWhinney (1987). Specifically, in the current model, as the sentence is being processed word by word, open class words are stored, and their referents are decoded based on the learned word-to-referent meanings. In parallel, closed class words and morphemes are collectively used to retrieve the associated sentence-to-meaning mapping. The constellation of closed class elements for a given grammatical construction type then serves as an index by which the sentence-to-meaning mapping template can be learned and retrieved. The referent items in the current sentence then fill in their respective thematic

roles based on the retrieved mapping, and the sentence is ‘understood.’ Details and examples can be found in [Dominey \(2000, 2003\)](#).

In a perfect joint attention condition, when the listener hears a sentence, the meaning that she currently entertains will be the one intended by the speaker. Referential uncertainty corresponds to those cases where the speaker’s and listener’s meanings are different. Part of the objective of this theoretical work was to study the interaction between knowledge of word meaning, and knowledge of grammatical constructions, to determine if the interaction between the two could further contribute to elimination of the indeterminacy problem at the level of sentences and their meanings.

Indeterminacy or referential ambiguity could be introduced to the system in the form of noise in the sentence, meaning (and word-referent) pairs ([Dominey, 2000](#)). This allowed us to determine the effects of this noise both on word learning and grammatical construction learning, and whether the mutual interaction between these two types of knowledge could overcome the referential ambiguity introduced by noise. Interestingly, the simulation studies revealed that when there was a reciprocal cooperation between semantic and syntactic knowledge (bootstrapping), the system was significantly more robust to the noise-induced indeterminacy. Indeed, what the results suggested was that in the developmental progression, a small set of concrete words could be initially learned in an ‘asyntactic first pass’ ([Gillette et al., 1999](#)) and their meaning refined through the synergy between syntactic and semantic information. With this small lexicon established, the set of learned grammatical constructions could then expand via semantic bootstrapping, with the resulting syntactic–semantic synergy that would allow an explosion both in the lexicon and in the inventory of grammatical constructions, and a significant aid in the fight against referential uncertainty through the syntax–semantics interaction ([Dominey, 2000](#)).

It is noteworthy that in this context of learnability, [Gillette et al. \(1999\)](#) performed a series of experiments in which adult human subjects were required to identify words by inspecting the environment (video sequences of mother–child interactions) for their use under different conditions that manipulated the type of information available to the subjects. Part of the objective of the study was to investigate the general observation that noun-learning is superior to verb learning in early language development. One of the interesting conclusions of this study was that rather than noun vs. verb (i.e. lexical category), the important factor for learnability was the concreteness or imaginability of the word. In early learning, observability (i.e. concreteness or imaginability) provides the essential information for determining a word’s referent, and the less concrete, the less observable the word.

Interestingly this can be restated such that concreteness is inversely related to referential uncertainty. In [Dominey \(2000\)](#) we thus attempted to simulate the respective effects of lexical category (noun vs. verb) and imaginability or referential uncertainty on word learning. Imaginability was simulated such that noise was added to the ⟨word, referent⟩ input pairs inversely proportional to the imaginability. As observed by [Gillette et al. \(1999\)](#) we observed that the variance in learning was almost exclusively attributed to imaginability rather than lexical category ([Dominey, 2000](#)).

This simulation work demonstrated that a relatively simple learning algorithm could learn grammatical constructions and provide the basis for a subsequent generalized,

generative language capability (related to [Tomasello, 1999, 2003](#)), in a context in which the indeterminacy problem had been reduced by mechanisms of attentional regulation (CDS and JA) as described above. In this theoretical context, then, CDS and joint attention are two among multiple mechanisms of varying complexity, including prosodic, semantic and syntactic bootstrapping, that allow the learner to reduce referential uncertainty at multiple levels, as demonstrated in the reviewed simulation studies.

5. Discussion

The focusing of attention onto relevant characteristics of the speech signal via CDS, and onto pertinent aspects of meaning in the scene via joint attention thus appear to play an important role in reducing the indeterminacy problem in language acquisition. Clearly, however, this analysis should not ignore the important role of more ‘automatic’ processing in language acquisition. Extensive investigations by [Jusczyk, Eimas, Saffran and others](#) (see [Jusczyk, 1997](#); [Kuhl, 2000](#) for reviews), have demonstrated infants’ impressive ability to learn statistical properties of speech simply by listening to language. For example, the phonemes that are linked within words co-occur with higher frequency than phonemes at word boundaries, and are thus more likely to be perceived as units by infants. The detection and exploitation of these statistical properties of ambient language thus allows infants to find word candidates in running speech before they know the meanings of words. This extraction or segmentation of words from the continuous speech signal is a crucial step in establishing the mapping between language and meaning. [Shi et al. \(1999\)](#) have recently gone one step further, demonstrating that newborn infants are sensitive to categorical acoustic differences between closed-class grammatical function words and open class content words. This early categorization capability likely plays a crucial role in language acquisition as specified by [Bates and MacWhinney \(1987\)](#) and simulated by [Blanc et al. \(2004\)](#) and [Dominey \(2000\)](#).

In this context of automatic processing, [Kuhl \(2000\)](#) recently observed that “what is innate regarding language is not a universal grammar and phonetics, but innate biases and strategies that place constraints on perception and learning. They allow infants to recover from language input the rules by which people in their community communicate.”

It should be made clear, however, that while these types of automatic perceptual and statistical learning allow the infant to extract rules about the internal structure of the speech signal, they provide little access to rules concerning the mapping of the acoustic signal to meaning. Likewise, they do not engage controlled attentional processes. With CDS an element of controlled processing enters the interaction. There, modulation of specific words, word categories and structural properties begins to focus the child’s attention on these properties, thus aiding the child in establishing the mapping between words and their referents, and in establishing the structural organization of her language.

With JA the child enters into an increasingly non-passive domain. There the child and caregiver engage in a triadic interaction in which an event or object becomes their shared center of interest. Via this explicit control of attention, the referential indeterminacy problem is significantly reduced, as the item of reference is indeed the object of the joint attention.

This raises an interesting point concerning the relation between the structure of the regularities or rules to be learned, and the nature of the automatic or controlled processes that will perform the learning. From the perspective of the regular internal statistical (distributional and phonotactic) structure of speech, indeterminacy is not a real problem, and relatively simple learning mechanisms can extract these speech-internal regularities (Christiansen & Chater, 2001; Dominey & Ramus, 2000). Indeed, non-human primates have now been demonstrated to be sensitive to statistical and prosodic regularities in speech samples, similar to human infants (Hauser, Newport, & Aslin, 2001; Ramus, Hauser, Miller, Morris, & Mehler, 2000, respectively). As the relation between this speech-internal structure and meaning is introduced, however, the indeterminacy problem becomes apparent. In this domain then, more controlled processing is required in order to reduce the potential indeterminacy that might otherwise render language learning difficult if not impossible. In particular, the ability to exploit joint attention, and to interpret the intentions of others becomes crucial. Interestingly, several studies have demonstrated that non-human primates can follow gaze direction of their conspecifics and humans under different conditions (Emery, Lorincz, Perrett, Oram, & Baker, 1997; Ferrari, Kohler, Fogassi, & Gallese, 2000), though their ability to follow human gaze is not intact until adulthood (Ferrari et al., 2000). Likewise, great apes appear to use gaze and pointing to indicate objects that they desire out of their reach (Krause & Fouts, 1997). Interestingly, however, they fail to interpret the communicative intentions of others. Tomasello, Call, and Gluckman (1997) tested apes and human children in a task in which novel visual cues were used to indicate which of three containers contained a food reward. Human children quickly picked up on the communicative intention of the cues in order to retrieve the reward. The apes in contrast appeared to process the cues as simple perceptual indices that were to be learned in a statistical, trial and error manner. In other words, the children interpreted each communicative act as an intentional attempt to direct their attention to the appropriate container, whereas the apes failed in this type of interpretation. This suggests that an important species difference is related to the ability to interpret the communicative intentions of others that likely plays a crucial role in the development of language.

The resulting picture of language acquisition might then proceed in the following manner. At or before birth, the infant will begin extracting structural regularities from ambient speech, in a largely automatic process that will significantly contribute to structural perception of speech, including word segmentation. At the same time, categorical differences will begin to form, laying the foundations for lexical categorization between open and closed class words, and other finer distinctions. In parallel, the presence of CDS will emphasize these structural regularities, as described above, thus initiating the more controlled, active direction of the infant's attention to relevant aspects of the speech signal. Note as mentioned above, however, that even ADS is child-exploitable in that the prosodic and related cues are still present though in a possibly reduced form. This controlled processing is amplified in the triadic interactions that are induced by joint attention shared between the infant and the caregiver. This shared attentional control focus the two interlocutors' attention onto a common referent, and thus directly combats the problem of indeterminacy. In this context of reduced indeterminacy, the mapping of meaning onto the structural components of the speech signal is now achievable. Even in the absence of joint

attention directed towards the infant, the infant's ability to follow the attentional cues of adults speaking with each other will provide useful referential information that will significantly reduce the referential uncertainty. As the lexicon develops, in parallel the infant begins to establish the structural mapping between grammatical constructions and the structural aspects of the meaning to which they refer. This will begin with the most basic constructions. The interaction between this developing syntactic knowledge and semantic knowledge will allow a synergistic bootstrapping that provides the basis for rapid progress both in lexical and grammatical development, again via a process of reducing the possible degrees of freedom, or referential uncertainty. Again, a crucial aspect of success in language acquisition is the reduction of the multiple degrees of freedom (or referential uncertainty) between language and its referent. A continuum of mechanisms including acoustic modulation that can be amplified in CDS, the construction of triadic interactions via joint attention, and the synergistic interaction of knowledge of word meaning and grammatical constructions appears to contribute at different stages to the reduction of these degrees of freedom.

At the outset of this paper we briefly described the indeterminacy problem in which, for the language learner, the correspondence between the signal and the referent in the world is extremely underspecified. That is, since the sound sequence could be referring to any one of an infinity of aspects of the surrounding environment, the language learner is lost. We then reviewed data from studies of CDS, and from the manipulation of joint attention that demonstrate that the referential indeterminacy problem can be radically reduced, by focusing the attentional processing of the learner both on the appropriate aspects of the signal, and on the appropriate aspects of the external world. From this perspective, the problem of the poverty of the stimulus takes on a new status. Its presence was previously used as a principal motivation for the absolute necessity of an innate universal grammar capability. If the problem is at least partially resolved, then the requirement for a pre-specified universal grammar capability is reduced. In this context, learning scenarios such as those presented by [Goldberg \(1995\)](#) and [Tomasello \(1999, 2003\)](#) become highly plausible. In these scenarios, learning plays a much more central role in language acquisition. The infant develops an inventory of grammatical constructions as mappings from form to meaning ([Goldberg, 1995](#)). These constructions are initially rather fixed and specific, and later become generalized into a more abstract compositional form employed by the adult that has been well described elsewhere ([Tomasello, 1999](#)).

It is beyond the scope of this paper to address the nativist vs. developmentalist debate. Nevertheless, the data reviewed here suggest that there are still issues to be considered in this context, particularly the interaction between learning and innate capabilities. In this context we can consider that CDS emphasizes regularities that the perceptual system is innately sensitive to and that have become exploited in human languages as information channels in the acoustic signal. Joint attention appears also to be an innate capacity, but is used in language acquisition as a non-language specific learning-related mechanism. The use of these innate, but non-language specific capabilities reduces the complexity of the acquisition problem, and thus reduces the functional requirements on the innate language-related learning mechanisms.

Acknowledgements

This work was supported by the Human Frontiers Science Program, including a Post-Doctoral Fellowship to CD, by the ACI for Integrative and Computational Neuroscience (Paris), and by the OHLL and Eurocores OMLL Initiatives. We appreciate constructive comments from Anne Reboul, and three anonymous reviewers.

References

- Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax* (pp. 117–134). Mahwah, NJ: Erlbaum.
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62(5), 875–890.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, 72(3), 708–717. Related Articles, Links.
- Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, 68(1), 48–57.
- Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney, & E. Bates (Eds.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, 14(2), 303–309.
- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech a window on early phonological development. In B. Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289–304). Boston: Kluwer Academic Publisher.
- Blanc, J. M., Dodane, C., & Dominey, P. F. (2004). Temporal processing for syntax acquisition: A simulation study. In *Proceedings of the 25th annual meeting of the cognitive science society, Boston*, in press.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33–B44.
- Broen, P. A. (1972). *The verbal environment of the language-learning child. American speech and hearing association monographs, number 17*, Washington, DC: American Speech and Hearing Association.
- Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.
- Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2, 1–20.
- Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding prior intentions enables two-year-olds to imitatively learn a complex task. *Child Development*, 73(5), 1431–1441.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63, 1–14.
- Carpenter, M., & Tomasello, M. (2000). Joint attention, cultural learning and language acquisition: Implications for children with autism. In A. M. Wetherby, & B. M. Prizant (Eds.), *Autism spectrum disorders: A transactional perspective (Vol. 9)* (pp. 30–54). *Communicative and language intervention series*, Baltimore: Paul H. Brooks Publishing.
- Chang, H. W., & Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child Development*, 48, 1666–1670.
- Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Christiansen, M., & Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Science*, 5(2), 82–88.

- Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, A. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*, 1570–1580.
- Cooper, R. B., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge: Harvard University Press.
- Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Oxford: Blackwell.
- Dawson, G., Munson, J., Estes, A., Osterling, J., McPartland, J., Toth, K., Carver, L., & Abbot, R. (2002). Neurocognitive function and joint attention ability in young children with autism spectrum disorder versus developmental delay. *Child Development*, *73*(2), 345–358.
- De Boer, B. G., & Kuhl, P. K. (2001). Infant-directed vowels are easier to learn for a computer model. *Journal of the Acoustical Society of America*, *110*(5 pt 2), 2703.
- De Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four language. *Journal of Child Language*, *67*, 297–319.
- Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Journal of Acoustic Society of America*, *76*, 57–66.
- Dominey, P. F. (2000). Conceptual grounding in simulation studies of language acquisition. *Evolution of Communication*, *4*(1), 57–85.
- Dominey, P. F., & Ramus, F. (2000). Neural network processing of natural language: I. Sensitivity to serial, temporal and abstract structure of language in the infant. *Language and Cognitive Processes*, *15*(1), 87–127.
- Dominey, P. F. (2003). Learning grammatical constructions in a miniature language from narrated video events. *Proceedings of the 25th annual meeting of the cognitive science society, Boston*, in press. (<http://www.isc.cnrs.fr/wp/DomineyISC-WP0.pdf>)
- Dominey, P. F., Hoen, M., Lelekov, T., & Blanc, J. M. (2003). Neurological basis of language in sequential cognition: evidence from simulation, aphasia and ERP studies. *Brain and Language*, in press.
- Dunham, P. J., Dunham, F., & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology*, *29*(5), 827–831.
- Eimas, P. D. (1975). Auditory and phonetics coding of the cues for speech: discrimination of the $r - 1$ distinction by young infant. *Perception and Psychophysics*, *18*, 341–347.
- Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., & Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (*Macaca mulatta*). *Journal of Computational Psychology*, *111*(3), 286–293.
- Falk, D. (2003). Prelinguistic evolution in early hominins: whence motherese? *Behavioral and Brain Sciences*, in press.
- Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*, 103–114.
- Fernald, A. (1976). The mother's speech to the newborn. Paper presented at the *Max-Planck Institute for Psychiatry, Munich*.
- Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, R. Golinkoff, M. Greenberg, C. Harding, & N. Bohannon (Eds.), *The origins and growth of communication* (pp. 5–29). Norwood, NJ: Ablex Publishing Corporation.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.
- Fernald, A. (1989). Intonation and communication intent in mother's speech to infants: is the melody the message? *Child Development*, *60*, 1497–1510.
- Fernald, A., & Kuhl, P. (1981). Fundamental frequency as an acoustic determinant of infant preference for Motherese. Paper presented at the *Biennial meeting of the society for research in child development, Boston*.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*, 279–293.
- Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants mothers' speech to new-borns. *Developmental Psychology*, *27*, 209–221.
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113.

- Fernald, A., Taeschner, T., Dunne, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501.
- Ferrari, P. F., Kohler, E., Fogassi, L., & Gallese, V. (2000). The ability to follow eye gaze and its emergence during development in macaque monkeys. *Proceedings of the National Academic Science, USA*, 97(25), 13997–14002.
- Fisher, C. L., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. *Child Development*, 67, 3192–3218.
- Gerken, L. A. (1996). Sentential processes in early child language: Evidence from the perception and production of function morpheme. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax* (pp. 411–425). Mahwah, NJ: Erlbaum.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition*, 73, 135–151.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.
- Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey (Vol. II)* (pp. 150–193). *Language: Psychological and biological aspects*. Cambridge: Cambridge University Press.
- Gleitman, L. R., & Wanner, E. (1982). *Language acquisition: The state of the art*. Cambridge: Cambridge University Press.
- Goldberg, A. (1995). *Constructions*. Chicago/London: University Chicago Press.
- Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British Journal of Developmental Psychology*, 4(3), 261–268.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 156–157.
- 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–B64.
- Hirsh-Pasek, K., & Golinkof, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Boston: MIT Press.
- Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young children. *Cognition*, 26, 269–286.
- 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, 548–567.
- Jusczyk, P. W. (1989). Perception of cues to clausal units in native and non-native languages. Paper presented at the *Biennial meeting of the society for research in child development, Kansas City*.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge: MIT.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23.
- Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402–420.
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.
- Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast of multisyllabic utterances by two-month-old infants. *Perception and Psychophysics*, 23, 105–109.
- Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merill-Palmer Quarterly*, 11, 95–122.
- Kaplan, P. S., Bachorowski, J. A. A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Sciences*, 13(3), 268–271.
- Kaplan, P. S., Bachorowski, J. A. A., Smoski, M. J., & Zinser, M. (2001). Role of clinical agnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy*, 2, 537–548.
- Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language*, 27, 343–358.

- Kemler Nelson, D. G., Hirsh-Pasek, K., Jusezyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 53–68.
- Klatt, D. K., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen, & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 69–86). Berlin: Springer.
- Krause, M. A., & Fouts, R. S. (1997). Chimpanzee (Pan troglodytes) pointing: hand shapes, accuracy, and the role of eye gaze. *Journal of Comparative Psychology*, 111(4), 330–336. Related Articles, Links.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22), 11850–11857.
- Kuhl, P. K., Andruski, J. E., Chistovitch, I. A., Chistovitch, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic in language addressed to infants. *Science*, 277, 684–686.
- Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics*, 31, 279–292.
- Langacker, R. (1991). *Foundations of cognitive grammar (Vol. 2). Practical applications*, Stanford: Stanford University Press.
- Lederer, A., & Kelly, M. H. (1991). Prosodic correlates to the adjunct/complement distinction in motherese. *Papers and Reports on Child Language Development*, 30, 55–63.
- Leslie, A. M. (1994). Toby and agency: Core architecture and domain specificity. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge: Cambridge University Press.
- Luce, P. A., & Charles-Luce, J. (1983). Contextual effects on the consonant/vowel ratio in speech production. Paper presented at the *Meeting of the acoustical society of america, Cincinnati*.
- Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers-to-children. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), (Vol. II) (pp. 173–184). *Child phonology: Perception*, New York: Academic Press.
- Mattys, S. L., Jusczyk, P. A., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38(4), 465–494.
- Menyuk, P. (1977). *Language and maturation*. Cambridge: MIT Press.
- Menyuk, P. (1994). Word acquisition: the important role of prosody. *The Study of Sounds*, 23, 218–222.
- Monnot, M. (1999). Function of infant-directed speech. *Human Nature*, 10, 415–443.
- Moore, C., & Dunham, P. (1995). *Joint attention: Its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000a). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, 21(3), 283–298.
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000b). Gaze following, temperament, and language development in 6-month-olds: a replication and extension. *Infant Behavior and Development*, 23(2), 231–236.
- Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge: MIT Press.
- Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.
- Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 20, 67–85.
- Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax* (pp. 263–281). Mahwah, NJ: Erlbaum.
- Morrongiello, B. A. (1984). Auditory temporal pattern perception in 6- and 12-month-old infants. *Developmental Psychology*, 20, 441–448.
- Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 13, 477–492.

- Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, 72(3), 718–735.
- Mundy, P. (1995). Joint attention and social-emotional approach behavior in children with autism. *Development and Psychopathology*, 7, 63–82.
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, 20(1), 115–128.
- Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: the contribution of nonverbal communication measures. *Journal of Child Psychology and Psychiatry*, 27, 657–669.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, 24(3), 1–11.
- Newport, E. L., Gleitman, H., & Gleitman, C. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow, & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–149). Cambridge, UK: Cambridge University Press.
- Nowik-Stern, A., Clarkson, M. G., Morris, M. K., & Bakeman, R. (1998). The effect of premature infants' speech preferences on mother–preterm interaction. *Poster, 11th Biennial international conference on infant studies, Atlanta, Georgia*.
- Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home video-tapes. *Journal of Autism and Developmental Disorders*, 24, 247–257.
- Papousek, M. (1995). Le comportement parental intuitif, source cachée de la stimulation musicale dans la petite enfance. In I. Deliège, & J. Sloboda (Eds.), *Naissance et développement du sens musical* (pp. 101–130). Paris: PUF.
- Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, 13, 539–545.
- Papousek, M., & Papousek, H. (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In L. Lipsitt, & C. Rovee-Collier (Eds.), *Advances in infancy research (Vol. 1)* (pp. 163–224). Norwood: Ablex.
- Papousek, H., & Papousek, M. (1987). Intuitive parenting: a dialectic counterpart to the infant's integrative competence. In: *Handbook of infant development* (pp. 669–720). New York: Wiley.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge: Harvard University Press.
- Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pinker, S. (1994). *The language instinct*. Paris: Odile Jacob.
- Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, 13, 85–100.
- Quine, W. (1960). *Word and object*. Cambridge, MA: Harvard University Press.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 280–281.
- Read, C., & Schreiber, P. (1982). Why short subjects are hard to find? In L. Gleitman, & E. Wanner (Eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge University Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33, 3–11.
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265–266.
- Selkirk, E. (1996). The prosodic structures of function words. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213).
- Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, 25, 169–201.
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11–B21.
- Shriberg, E., Stolcke, A., Hakkani-Tür, D., & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32, 127–154.

- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39–91.
- Spring, D., & Dale, P. (1977). The discrimination of linguistic stress in early infancy. *Journal of Speech and Hearing Research*, 20, 224–231.
- Steels, L., & Kaplan, F. (2002). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge: Cambridge University Press.
- Stern, D. N., Spieker, S., Barnett, R. K., & McKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *Journal of Child Language*, 10, 1–15.
- Stern, D. N., Spiker, S. B., & McKain, K. (1982). Intonation as signals in maternal speech to pre-linguistic infants. *Developmental Psychology*, 18, 727–735.
- Tomasello, M. (1988). The role of joint attention in early language development. *Language Sciences*, 11, 69–88.
- Tomasello, M. (1992). The social bases of language acquisition. *Social Development*, 1, 67–87.
- Tomasello, M. (1999). The item-based nature of children's early syntactic development. *Trends in Cognitive Science*, 4(4), 156–163.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Tomasello, M., Call, J., & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development*, 68(6), 1067–1080.
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4, 197–212.
- Vogt, P., & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, 6(1) (<http://jasss.soc.surrey.ac.uk/6/1/4.html>).
- Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, 43, 230–246.