

Activity types and child-directed speech: a comparison between French, Tunisian Arabic and English

LUDIVINE GLAS

Laboratoire Dynamique du Langage (UMR5596, Université Lyon2)

ludivine.glas@wanadoo.fr

CAROLINE ROSSI

*Institut des Langues et des Cultures d'Europe, Amérique, Afrique,
Asie et Australie (EA7356, Université Grenoble-Alpes)*

caroline.rossi@univ-grenoble-alpes.fr

RIM HAMDI-SULTAN

Laboratoire Savoirs, Textes, Langage (UMR 8163, Université Lille3)

rim.hamdi@univ-lille.fr

CÉDRIC BATAILLER

*Laboratoire de Psychologie et NeuroCognition (UMR5105, Université
Grenoble-Alpes)*

cedric.batailler@univ-grenoble-alpes.fr

and

HACENE BELLEMMOUCHE

Laboratoire LIRADDI, Université d'Alger 2

hacenebl@yahoo.fr

Abstract

Quantity and quality of input affect language development, but input features also depend on the context of language emission. Previous research has described mother-child interactions and their impact on language development according to activity types like mealtimes, book

We would like to thank Darine Saïdi from the University of Paris 8, for her help in understanding Tunisian data. This study was conducted within the PREMS project (PI S.Kern), funded by the French National Research Agency (ANR-11-BSH2-0009) <<http://www.agence-nationale-recherche.fr/Projet-ANR-11-BSH2-0009>>

reading, and free play. Nevertheless, few studies have sought to quantify activity types in naturalistic datasets including less-studied languages and cultures. Our research questions are the following: we ask whether regularities emerge in the distribution of activity types across languages and recordings, and whether activities have an impact on mothers' linguistic productions. We analyse input for two children per language, at three developmental levels. We distinguish three activity types: solitary, social and maintenance activities, and measure mothers' linguistic productions within each type. Video-recorded activities differ across families and developmental levels. Linguistic features of child-directed speech (CDS) also vary across activities – notably for measures of diversity and complexity – which points to complex interactions between activity and language.

Keywords: Child-directed speech, crosslinguistic comparison, activity context, natural video recording, longitudinal study

Résumé

La quantité et la qualité de l'input ont un impact sur le développement du langage, mais certaines caractéristiques de l'input dépendent aussi du contexte situationnel, autrement dit du type d'activité, dans lequel les énoncés sont produits. Certains travaux ont décrit l'impact des interactions mère-enfant sur le développement du langage en fonction du type d'activité comme les repas, la lecture ou le jeu libre. Cependant, très peu de travaux ont cherché à quantifier la répartition des différents types d'activités dans des suivis longitudinaux, en intégrant des langues moins souvent étudiées. Nos questions de recherche sont les suivantes : le codage par types d'activités révélera-t-il des régularités indépendantes des langues et des enregistrements, et les types d'activités observés ont-ils un impact sur le langage adressé à l'enfant (LAE) ? Les enregistrements analysés concernent l'input de deux enfants par langue (tunisien, anglais américain et français), à trois étapes développementales. Nous distinguons trois types d'activités : les activités sociales, solitaires et de maintenance, et les productions des mères sont mesurées au sein de chaque type. Les activités enregistrées varient d'une famille à l'autre, et à chaque étape développementale. Des variations du LAE sont également observées d'un type d'activité à l'autre : elles concernent la diversité et la complexité mesurées, et suggèrent qu'il n'existe pas de lien univoque entre type d'activité et caractéristique de l'input.

Mots-clés: Langage adressé à l'enfant, comparaison inter-langue, contexte de production, suivis longitudinaux, enregistrements naturalistes

1. INTRODUCTION

Cross-linguistic comparisons have shown similarities across children's expressive vocabularies (e.g., Gopnik and Choi 1990, Jackson-Maldonado et al. 1993, Caselli et al. 1995, Thordardottir and Weismer 1996, Caselli et al. 1999, Hamilton et al. 2000, Maital et al. 2000). In fact, as they acquire their mother tongue, all children go through the same developmental levels: vocabulary size and age of acquisition have been shown to be very similar for French and English (Poulin-Dubois et al. 1995), but also in other Romance languages such as Mexican Spanish (Jackson-Maldonado et al. 1993), and Italian (Caselli et al. 1995). Two different levels are generally acknowledged across languages. These levels are characterised by different rhythms in new word acquisition and by the use of words belonging to different

grammatical categories. On the whole, despite crosslinguistic or inter-individual differences, the same reorganisations in children's vocabulary occur with regard to word class between 12 and 36 months of age. A first level is characterised by rather slow acquisition of words that are mostly nouns (Jackson-Maldonado et al. 1993, Caselli et al. 1995, Bassano 1998, Maital et al. 2000, Kauschke and Hofmeister 2002, Bornstein et al. 2004), followed by vocabulary spurt and first word combinations. Nevertheless, variability among children has been argued to be a result of various factors, including gender (girls having slightly larger productive vocabularies, as shown by Eriksson et al. (2012) using adapted Communicative Development Inventories (CDIs) for 10 non-English language communities) birth rank (e.g., Bates et al. 1991, Fenson et al. 1994, Maital et al. 2000) and differences in input. Input may vary both qualitatively and quantitatively as a result of socio-economic status (SES), the type of the language being acquired and the interactions or activity types. The impact of endogenous factors as well as SES has been controlled for in our study, but will not be dealt with in this paper, since our interest here lies in input differences. In what follows, we discuss studies that look at parents' behaviour and input as possible explanations for variability in children's language.

1.1 Input and linguistic environment

Input is crucial to language acquisition. Quantity and quality of input have been singled out as predictors of language development, from the very first levels, as well as in later levels of development (Huttenlocher et al. 1991, Hart and Risley 1995, Bornstein et al. 1998, Florin 1999, Weizman and Snow 2001). From the point of view of quantity, various studies have indeed shown the impact of frequency on the order of acquisition of lexical items (Huttenlocher et al. 1991, Goodman et al. 2008). Thus, in 16-month-old children, order of acquisition is closely linked with the relative frequency of words acquired in parents' speech (Huttenlocher et al. 1991). Overall, the words and grammatical categories heard the most often should be learned earlier by children (Choi and Gopnik 1995, Tardif 1996, Goodman et al. 2008). Together with quantity, quality is a possible source of variation which has been analysed in caregiver-child interaction, for example, based on mothers' responsiveness and volubility (Vanormelingen and Gillis 2016) or on the pragmatic values of caregiver utterances (see Farran and Haskins 1980, Hoff 2006, Rowe 2008 on the impact of directives versus conversational utterances). Looking at the children's linguistic environment, studies have shown that cross-linguistic differences affecting morphology, salience, as well as frequency and pragmatic aspects had an impact on lexical development, especially on vocabulary composition. For example, a noun bias in expressive vocabulary is observed in children acquiring English (e.g., Au et al. 1994, Bates et al. 1994, Fenson et al. 1994, Tardif et al. 1997, Goldfield 2000), Italian (e.g., Caselli et al. 1995, Tardif et al. 1997), Spanish (Jackson-Maldonado et al. 1993), French (Poulin-Dubois et al. 1995, Bassano 2000, Parrisé and Le Normand 2000), and Hebrew (Maital et al. 2000), but these results could not be replicated in other languages like Korean and Mandarin, where conflicting evidence has been found (see Gentner 1982, versus Au et al. 1994). One way of

accounting for observed differences has been to look at pragmatic aspects of language. Different linguistic and cultural communities also differ as regards the focus of caregivers' discourse to children, emphasizing different environmental and linguistic aspects during caregiver-child interaction (Bornstein et al. 1992). For example, American mothers tend to focus on objects and request object labels from their children for a variety of socio-cultural reasons: they have been shown to focus more on objects than Chinese and Japanese mothers (Tamis-LeMonda et al. 1992, Fernald and Morikawa 1993, Gopnik et al. 1996). More recently, variation has been shown to depend more on context differences than on language typology: in Altinkamış et al. (2014), nouns prevailed in both French and Turkish child-directed speech in book-reading contexts, and there were more verbs in toy play. In an effort to "put the noun bias in context" (Tardif et al. 1999), such studies show an effect of context on linguistic measures, with possible impacts on child language acquisition. A different approach, going back to Bruner's (1981) analyses, consists in starting from context and analysing how it relates to language acquisition.

1.2 Influence of context on interactions and input

In the wake of Vygotsky's social interactional conception of language, Bruner insists on social context for language acquisition, underlining the importance of what he called formats in social interaction and showing that they are a crucial level in language development (Bruner 1981). Bruner's seminal work has certainly influenced methods and topics in the field, but to our knowledge, research linking context with language acquisition is still relatively scarce, more focused on specific contexts like meals or book reading, and mostly aimed at demonstrating SES- or community-related differences (e.g., the part played by parents' free time and dedication to children's development – and notably time spent reading (Weizman and Snow 2001) – and the impact of engagement and stability of mother-child dyads on later development (Leyendecker et al. 1997b)). Such studies outline input characteristics and link this input to child language development, for example for mealtime context (Ely et al. 2001, Snow and Beals 2006), book reading (Reese and Cox 1999, Choi 2000, Raikes et al. 2006), or free play (Choi 2000, Newland et al. 2001). The variety of contexts (Hoff-Ginsberg 1991) as well as duration of caregiver-child interaction (Snow et al. 1982, Hoff et al. 2002) have been shown to have an impact. This line of research has also related certain kinds of more specific activities to more advanced language abilities like literacy. The next paragraph is a summary of the general contribution of previous studies to our understanding of the part played by social interactional contexts in language development. In line with Bruner's work, play activities are often related to joint attention (e.g., Newland et al. 2001) and more generally to fundamental pragmatic characteristics of interaction. Interactions in play contexts have also been used to assess the quality, stability, engagement and sensitivity of caretakers interacting with their children (interactional style) (Masur and Gleason 1980, Tamis-LeMonda et al. 1996, Leyendecker et al. 1997a, Leyendecker et al. 1997b, Newland et al. 2001, Yont et al. 2003, Tamis-LeMonda et al. 2004). Although SES-related differences are generally observed

across contexts, book-reading contexts are a notable exception, in which both working-class and middle-class mothers have been shown to interact with their children in supportive ways (Snow et al. 1976, Lewis and Gregory 1987, Wiley et al. 1989, Hoff-Ginsberg 1991). Indeed, because mothers use more referential language while reading (Raikes et al. 2006), and because of the more complex nature of child-directed speech in such contexts (from the point of view of both lexical and syntactic diversity), large-scale studies have shown a positive impact of the time spent reading (Weizman and Snow 2001) with no clear SES-related differences (Hindman et al. 2014). Another line of research sought to go against the bias towards contexts “defined by researchers – usually toy play and book reading” (Hoff-Ginsberg 1991: 782) by looking at interactions in contexts such as mealtime or other goal-directed caretaking interactions, such as dressing or household chores, in order to capture “most children’s typical experiences” (Hoff-Ginsberg 1991: 782). Such contexts have been shown to differ from more usually studied contexts, in that they had smaller rates of child-directed speech and higher rates of conversation-eliciting utterances, together with lower lexical diversity (Hoff-Ginsberg 1991). The above-mentioned studies generally link the behaviour of dyads at a given time to later language and cognitive development of children. A different line of research has focused on a given linguistic feature of language development, for example the nouns bias (Choi 2000, Kern et al. 2012, Altunkamış et al. 2014), or on mothers’ conversational style (Bornstein et al. 1992, Haden and Fivush 1996, Kloth et al. 1998, Flynn and Masur 2007, Golinkoff et al. 2015, Kelly et al. 2015) in order to track variation or show stability across contexts. Such studies have shown, for instance, that mothers used more action-oriented utterances and hence more verbs in toy-play contexts than in book-reading contexts (Kern et al. 2012). A final line of research has taken activities into account with a view to contextualising vocabulary development within the Human Speechome corpus (D. Roy et al. 2006a, 2006b; D. Roy 2009; B. C. Roy et al. 2009, 2012; B. C. Roy et al. 2015). Roy et al. have contributed to operationalising Bruner’s format by establishing an exhaustive list of activities with dense and longitudinal data about one child. The idea behind the Speechome corpus is to gain insight into the influence of family environment on language acquisition, thus taking both linguistic and non-linguistic behaviours into account. B. C. Roy (2014) has explored activities in terms of spatial and temporal distribution as well as word use. Having built a model of context distributions and mapped it on every individual word in the corpus, he was able to single out more or less distinctive words along that dimension. The main finding is that “words used in distinctive spatial, temporal, and linguistic contexts are produced earlier, suggesting they are easier to learn” (B. C. Roy et al. 2015: 1). B. C. Roy et al.’s ground-breaking operationalisation of activity contexts is, to our knowledge, unparalleled in the literature. However, it remains to be seen how their methods could be applied to more traditional longitudinal data, which are still widely used in language acquisition studies. In the present study we use activity contexts in longitudinal data for six children, across three different languages. Our coding system and analyses are presented here as a possible application of B. C. Roy et al.’s methods: we discuss results with a view to assessing what could be achieved if it were to be used on more data.

1.3 Our study

As shown in our analysis of previous literature, the notion of the context where acquisition takes place has not often been analysed as such. And yet, we know that context has an impact: according to the type of social activity, linguistic input changes in terms of quantitative, qualitative and pragmatic vocabulary. Inspired by recent research on activity contexts, the aim of the present exploratory study is twofold. Our first goal is to observe which kinds of activities occur in longitudinal video recordings of different families, languages and cultures: our study thus focuses on three developmental levels and three different languages. Second, we would like to see if activity types may be linked with variations in child-directed speech. In order to address these goals, we made two hypotheses. According to our first hypothesis, while different patterns of activities will be found in each recording, regularities may emerge if activities are categorised into a limited number of activity types. According to our second hypothesis, differences in child-directed speech may be observed as a function of activity types.

2. METHOD

We used longitudinal recordings of children from six different families and three different languages and cultures to test our hypotheses. Rather than focusing on linguistic features like morphology and salience, or social features like birth order, this article starts from an analysis of the context of speech productions and seeks to show whether and how it relates to linguistic factors like input frequency, diversity or complexity.

2.1 Corpus, population, and description

Our data consist of longitudinal recordings from three distinct linguistic communities: one Romance language (French) one West Germanic language (American English) and one Semitic language standing outside existing classifications, being non Indo-European (Tunisian Arabic). As shown in a recent paper (Kelly et al. 2015), while the literature is very dense about language development in some linguistic communities, investigation of some less-documented languages remains a major challenge. Thus, there are very few studies on the acquisition of Arabic, and even fewer studies on children acquiring dialectal Arabic (Badry Zalami 2006). When such studies exist, they deal with either children's learning of formal and so-called "standard Arabic", which is a later acquisition, or with the acquisition of certain varieties like the Egyptian Arabic dialect (Omar 2007) or the Jordanian dialect (Amayreh 2003).

The same recording procedure was followed for all three corpora, namely one-hour recordings at 15-day intervals at the child's home. The instructions given were always the same: behave exactly as you would if there was no observer (Pellegrini et al. 2004). Our three datasets are taken from the Providence corpus (Demuth et al. 2006) for American English, from the PREMS corpus (Kern et al. 2009) for Tunisian Arabic and from the French OHLL corpus (Kern 2005). We used data from two children per language: one boy and one girl. All families have a high SES. Birth rank was not controlled, but the information was collected for

| Language | Gender | Birth order | Age at linguistic level | | |
|------------------|--------|-------------|-------------------------|--------------|--------------|
| | | | 10 words | 50 words | 100 words |
| American English | F | 2 | 14;00 | 18;18 | 20;22 |
| | M | 4 | 16;28 | 19;28 | 21;08 |
| French | F | 1 | 15;26 | 19;29 | 21;20 |
| | M | 2 | 14;29 | 18;14 | 19;4 |
| Tunisian Arabic | F | 2 | 17;1 | 21;3 | 22;21 |
| | M | 1 | 14;16 | 21;22 | 25;6 |
| Average | | | 15.56 | 19.97 | 21.78 |

Table 1: Corpus description.

each child and is set out in Table 1. Participants are not shown in the table, but they were coded as headers in each transcription and taken into account in our analyses.

The children were recorded before the first word and until they produced 200 different words, but for the purpose of our fine-grained analyses of activity contexts we used data at only three key linguistic levels, that is to say 10, 50 and 100 words produced. Each session was 50 minutes long. The 10-word level was the most homogeneous early level in our data, and it corresponds to a well-described milestone in language development (Tardif et al. 2008). The same is true of the 50- and 100-word levels, which precede and follow the lexical spurt (Bloom 1973, Benedict 1979, Huttenlocher et al. 1991, Fenson et al. 1993, Kern 2010). Thus, we worked on 2.5 hours of recording for each child and 5 hours for each linguistic level, amounting to a total of 15 hours of recording.

In order to obtain equivalent linguistic levels while maximising comparability across datasets, we calculated cumulated types across all recordings, so that each linguistic level takes into account every word produced by the child up to that point. As can be seen in Table 2, where both raw and cumulated scores appear, for each session chosen for analysis we have rather homogeneous results across children and languages.

As expected, when looking at those measures no child stands out across all levels. Nevertheless, for the 100-word level girls systematically score above boys in word type in all three linguistic groups: 99 versus 37 for US children, 59 versus 32 for French children and 99 versus 50 for Tunisian children.

We computed mean length of utterance (MLU) from the second level onward. More variation can be seen in MLU counts, which are not necessarily a function of lexical diversity. In our data, they are probably more of an index of the children's conversational style. Indeed, the French girl's longer MLU is linked with repeated utterances based on the same chunks.¹ As a result, the number of types remains relatively low. The MLU for the Tunisian children does not change as quickly, but this is probably the result of *Phon* transcriptions of this Semitic language (Omar 2007: 9): each transcribed word includes affixes and may correspond to several words in

¹For instance: [est qui ça ?] appears 24 times in the last level.

| Language/ Gender ¹ | Age | Total utterances | Meaningful utterances | Utterances containing words | Cumul. types | Total types | Total tokens | MLU |
|-------------------------------|-------|------------------|-----------------------|-----------------------------|--------------|-------------|--------------|------|
| 10-word level | | | | | | | | |
| AE-M | 16;28 | 141 | 20 | 18 | 7 | 7 | 18 | |
| AE-F | 14;00 | 88 | 8 | 6 | 7 | 7 | 10 | |
| Fr-M | 14;29 | 418 | 17 | 14 | 6 | 6 | 15 | |
| Fr-F | 15;26 | 604 | 22 | 16 | 4 | 4 | 16 | |
| TA-M | 14;16 | n/a | 13 | 13 | 10 | 10 | 15 | |
| TA-F | 17;1 | n/a | 47 | 45 | 30 | 30 | 49 | |
| Mean | | 312.75 | 21.17 | 18.67 | 10.67 | 10.67 | 20.5 | |
| StDev | | 242.15 | 13.61 | 13.53 | 9.64 | 9.67 | 14.21 | |
| 50-word level | | | | | | | | |
| AE-M | 16;28 | 282 | 95 | 95 | 40 | 35 | 157 | 1.9 |
| AE-F | 14;00 | 317 | 76 | 76 | 32 | 27 | 121 | 1.87 |
| Fr-M | 14;29 | 487 | 112 | 92 | 33 | 28 | 99 | 1.36 |
| Fr-F | 15;26 | 375 | 156 | 105 | 30 | 27 | 145 | 1.49 |
| TA-M | 14;16 | n/a | 47 | 28 | 19 | 19 | 31 | na |
| TA-F | 17;1 | n/a | 142 | 136 | 86 | 86 | 172 | 1.27 |
| Mean | | 365.25 | 103.2 | 88.67 | 40 | 37 | 120.83 | 1.57 |
| StDev | | 89.77 | 45.44 | 35.78 | 23.54 | 24.54 | 51.1 | 0.29 |
| 100-word level | | | | | | | | |
| AE-M | 16;28 | 418 | 131 | 131 | 63 | 37 | 229 | 2 |
| AE-F | 14;00 | 357 | 162 | 158 | 123 | 99 | 247 | 1.96 |
| Fr-M | 14;29 | 847 | 199 | 151 | 73 | 54 | 165 | 1.31 |
| Fr-F | 15;26 | 307 | 218 | 186 | 73 | 59 | 335 | 2.28 |
| TA-M | 14;16 | n/a | 90 | 82 | 56 | 56 | 117 | 1.43 |
| TA-F | 17;1 | n/a | 206 | 179 | 99 | 99 | 232 | 1.3 |
| Mean | | 482.25 | 167.67 | 147.83 | 81.17 | 67.33 | 220.83 | 1.71 |
| StDev | | 247.37 | 49.79 | 37.84 | 25.16 | 25.7 | 74.58 | 0.42 |

Table 2: Children’s linguistic measures at each level

¹AE: American English; Fr: French; TA: Tunisian Arabic; M: male; F: female.

English or French. This had no impact on the comparability of our data, however, since types were retrieved and counted manually in order to establish cut-off points for the three developmental levels.

2.2 Transcription and coding

This section is mainly concerned with coding, but since the process of transcription required decisions that were likely to have an impact on our analyses (especially as regards the linguistic measures), we begin by explaining how we transcribed. We focused on child-directed speech only, since other utterances were hardly ever relevant to ongoing caregiver-child activities. Besides, quantity of child-directed speech is the best predictor of overall discourse quantity heard by a child (Weisleder and Fernald 2013). In addition, because only mothers' and children's utterances could be transcribed in our Tunisian dataset, we restricted our analyses to mother-child interactions in all three languages. This did not lead us to exclude much in the French and American data, especially since the observer was often the only other person present and was not supposed to take part in interactions.

2.2.1 Transcription

All main speakers in our recordings – that is, at least child and mother – were orthographically transcribed with *Phon* software (Rose et al. 2005) using CHILDES rules (MacWhinney 2009). A major problem in transcribing and analysing our data was to find a common definition for a word in all three languages. As a rule, we used the widely accepted definition by Vihman and McCune (1994). We considered as a word any linguistic form spontaneously produced, in an appropriate or relevant context, when this linguistic form was near or similar to the adult form. Besides, since we focused on lexical diversity, we included for instance, alphabet letters in our transcriptions and analyses.² Only unintelligible items (coded yyy in CHILDES datasets) were excluded. We assigned word limits based on written language (Vihman and McCune 1994).

2.2.2 Grammatical categories

We focused on two grammatical categories that previous research has shown to be relevant both cross-linguistically and across activity contexts: nouns and verbs. Each word in the corpus was coded as a noun, as a verb, or as neither, which we treated as a third generic category.

In order to obtain comparable data, we adopted the coding categories used in previous studies (Choi 2000). Thus, our noun category included proper names, kinship terms like *maman* and common nouns, but excluded pronouns. Our category of verbs included main verbs, but excluded auxiliaries and copulas. This was particularly helpful in Tunisian Arabic, since auxiliaries and main verbs often appeared as one word in our transcriptions. Besides, since our coding was done manually, it

²However, these were not counted when calculating cumulated types to determine our key levels.

enabled us to distinguish noun-verb homonyms, as illustrated by the two sentences read by the American boy's mother to her son in our data:

- (1) The count loves counting things Ernie loves to drum.
- (2) Here's a drum to bang here's a phone to ring.

Last but not least, no lemmatisation was performed on our data, because it could not be done for Tunisian Arabic. Thus, in what follows, we chose to present tokens rather than non-lemmatised types.

2.2.3 *Ethological coding*

Although the premise of our approach is original, the resulting categorisation of activities is not unique: previous research used equivalent taxonomies of child activity. For example in Leyendecker et al. (1997a) and Leyendecker et al. (1997b), activities were divided into five exclusive and exhaustive contexts: feeding, caretaking, toy play, social interaction and no interaction. Overall, functional and social contexts clearly had an effect on interactional experiences. SES effects on verbal and other interactional measures were limited to a subset of contexts and may thus represent the infants' overall experiences quite poorly. Such results show that comparisons based on a single context may be inadequate for studies of participants from differing socioeconomic backgrounds and point to the need for further explorations of contexts. In line with Bruner's (1981) notion of format and B. C. Roy et al.'s (2013) operationalisation of the notion as a core element in defining activity contexts, we coded the contexts of mother-child interactions. In contrast to what was done with Speechome data, however, we designed our own method and coding scheme. Using traditional tools of ethological measurement, we first drafted a map of activities occurring in our families on the entire French corpus, that is to say on 330 hours of video recording. Thus we coded for a wide variety of activities. Having done that, we distinguished three main activities that occur regularly in any child's day: exploration, maintenance, and social activities. As shown in Table 3, we consider as solitary activity all activities of environment exploration (e.g., looking out the window) and solitary play (e.g., playing alone with a toy). Maintenance activities include all activities revolving around food (dinners or snacks), health and hygiene (bath time, blowing the child's nose), while social activities include all activities of social play, book reading, and social speech (e.g., discussing previous experience).

We started by looking for regularities and variation across our three main types of activities but went back to our finer-grained subtypes where appropriate. We did not, however, code for activities that included only a few utterances and were nested in a previous activity which continued afterwards. Most of the time, those nested activities were failed attempts to change the focus of the interaction. For example, if while book reading a mother tries to give a glass of water to her child and the child refuses and continues to read so that book reading with mother starts again, we included all utterances in a book reading activity. Examples from our corpus are available in Appendix 1. Each of the first three authors of this paper was responsible for coding in one language, and there were regular meetings to discuss coding

| Activity type | Activity subtype | Example |
|---------------|--------------------------|---|
| Solitary | Environment exploration: | Play with washing machine |
| | Solitary play: | Manipulation game alone |
| Maintenance | Feeding: | Dinner, snacking, drink |
| | Cleanliness and hygiene: | Diaper change, bath, handwashing |
| Social | Social play: | Play tea time, memories play, hide-and-peek |
| | Book reading: | Story or picture book reading by caregiver |
| | Social speech: | Conversation with caregiver about the day, friend, pets ... |

Table 3: Coding categories of activities.

and to check reliability across languages. Time spent interacting or doing things without speaking is not taken into account and not coded, so we do not have measures of total time for activities in our recordings. Our measures include total number of utterances as well as global duration of activities including verbal productions by at least one participant.

2.3 Measures

Even if a child needs to hear a certain quantity of input before acquiring a word, raw frequency is probably not the best predictor for word learning; quality and interaction may be more accurate (Clark 2009, Cartmill et al. 2013). In what follows, we began with raw measures and then computed more qualitative measures such as lexical diversity and noun-verb proportion. We measured activity duration for each level and child. This enabled us, first, to see how activities were spread in the sessions we analysed, and second, to adjust lexical measures depending on durations. Additionally, we computed overall durations for each activity type over each recorded session. All linguistic measures were computed using CLAN software. At each linguistic level, for whole sessions, and depending on the activity when we had enough data, we calculated token and types of word, number of utterances and mean length of utterance as well as lexical diversity, using the D measure (Malvern et al. 2004) and the VOCED (Vocabulary Diversity) command in CLAN. Considering the importance of noun versus verb differences in the literature and the observed variations across activities, we also calculated the number of nouns, verbs, and other words in each recorded session, as well as for each activity. Proportions of nouns and verbs (tokens) per utterance were computed for each activity type, as was done in Choi's landmark study (Choi 2000). Using per-utterance ratios also helped us correct for the differences that resulted from our definition of word boundaries in Tunisian Arabic.

3. ANALYSES

Our study is aimed at assessing the impact of recorded activities on the linguistic measures that can be obtained based on longitudinal corpora. Because of the

relatively small size of our sample, it is unlikely that major differences linked with different genders or birth-ranks will be found. Thus, in what follows we will not take gender or birth-rank into account unless we notice differences that might be linked with those characteristics.

Besides, due to the complexity and exploratory character of data collection, coding and analysis, we did not have enough data to run powerful statistical tests. The relatively small size of our sample, which includes data from only six children, does not allow for any inferential statistics. Indeed, the probability that our tests will result in type I error is stronger than that of finding any effect. In order to avoid over-generalising from a small amount of data, we present only descriptive statistics that will be analysed and interpreted as such. We sought to produce a coherent overview of our data by using the same bar charts for all the measures presented in this section.

3.1 Duration of activities

In order to assess the relevance of our first hypothesis and track the presence or absence of regularities in the distribution of activities, we started from our three main types of activities and compared durations across languages, children and developmental levels.

3.1.2 Global duration

Across all languages and developmental levels, a large majority of coded activities are social activities: they represent more than half of all the coded data. Solitary activities make up about a quarter of the data, and maintenance activities cover an hour and a half only (Figure 1).

3.1.3 Detailed duration of activities

In spite of clear variation across datasets, Figure 2 shows that social activities (including play interaction, discussion, etc.) are the most stable across our coded sessions and children: they amount to 10.51 minutes minimum and 42.48 minutes maximum in our 50-minute recordings. Solitary activities are less important but present in nearly all sessions, amounting to a maximum of 35 minutes and a minimum of less than 1 minute per session. Finally, maintenance activities are relatively infrequent and by far the least represented category in our data. Maintenance activities have been found in only 9 out of 15 coded sessions, and they cover less than 5 minutes altogether, including mostly short activities that are linked with snacking or blowing the child's nose.

Beyond the sheer variety, visual patterns seem to appear vertically rather than horizontally, which points to variation across languages, and no clear impact of developmental levels on activity types, apart from the French dataset where an increase of social activities and a corresponding decrease of solitary activities are observed for both children. When we look at the overall recorded duration per children, pooling all sessions together, the main result is the quantity of social activity for all children. Duration of social activity never falls under 50 minutes, that is a third of the overall recording. The maintenance and solitary activities are more variable.

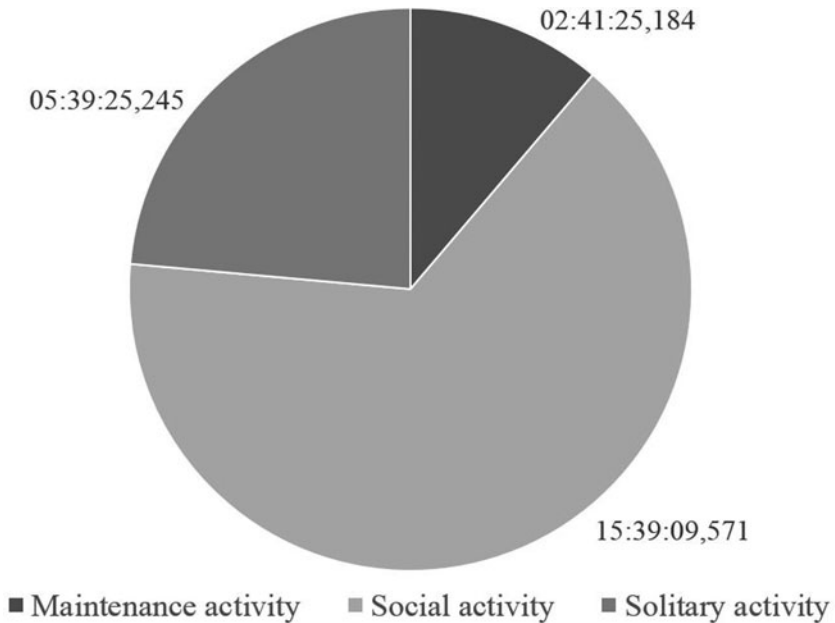


Figure 1: All corpus activities.

On the whole, rather than showing differences at each developmental level, or as a result of the different languages and cultures, the observed variability in recorded activities points to subtle differences in what is actually going on in the data, and it is worth noting that those differences are seldom mentioned in longitudinal studies of language development. Indeed, although recording methods are generally similar, and similar instructions are given to families – telling them to interact with their child as they usually do – recorded activities may differ considerably as a result of various external constraints, including the observer's presence. In our Tunisian data, there were many more child-observer interactions, and the child's siblings and cousins were more often around, so that social activities were predominant. These elements are crucial in order to understand the findings presented here. Therefore, more detailed elements are provided and discussed in section 4.

3.2 Common lexical measures: variation across activities

In order to assess the relevance of our second hypothesis, according to which activity types may have an impact on linguistic measures, we looked for differences in child-directed speech across our three main types of activities, again as a function of languages, children and developmental levels. The first element to be taken into account as regards child-directed speech is the existence of disparities in the

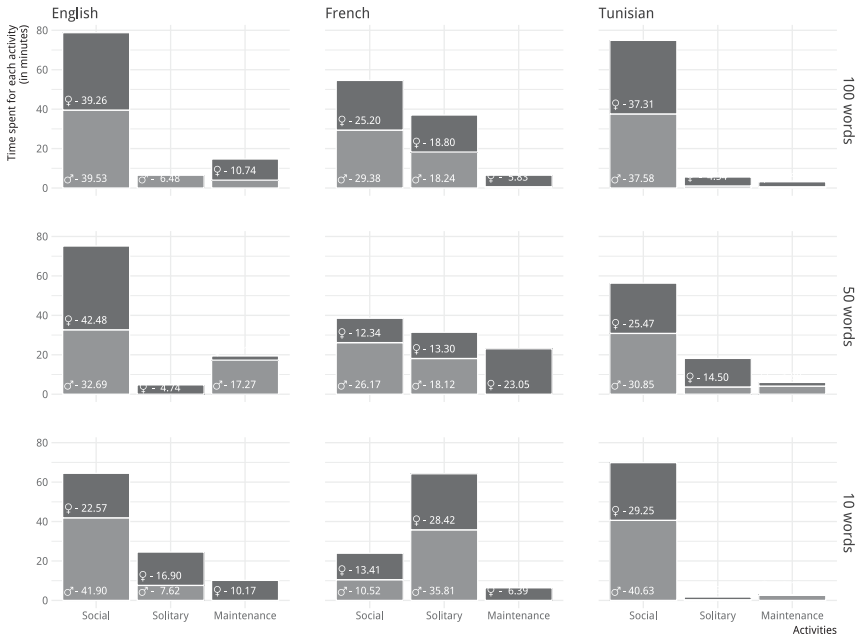


Figure 2: Activity duration per language, gender, and developmental level (labels corresponding to minutes spent).

amount of speech addressed to the children in our recordings. We used utterance and word counts to assess these differences.

3.2.1 Utterances

As shown in Figure 3, the total number of utterances directed to children in two and a half hours of recording varies a lot, with Tunisian Arabic child-directed speech clearly standing out as containing fewer utterances. This is partly due to the fact that the observer in Tunisian recordings interacted more with the child than was the case in the other two languages, and observer speech was not included in the present study.

The overall proportion of utterances in each activity context, language and level is by and large similar to the activity durations recorded in Figure 2, which suggests that verbal interactions are constant throughout the recordings, with only a few exceptions – notably, the small amount of speech addressed to the children in solitary contexts (with the exception of the French girl), which suggests that for some time at least, there is very little dialogue. The same trend is observed for our Tunisian data in social contexts. It is also worth noticing that in the American recordings only, utterances per minute become more numerous at each developmental level (especially in social context), so that mothers appear to be adjusting to their child’s linguistic development.

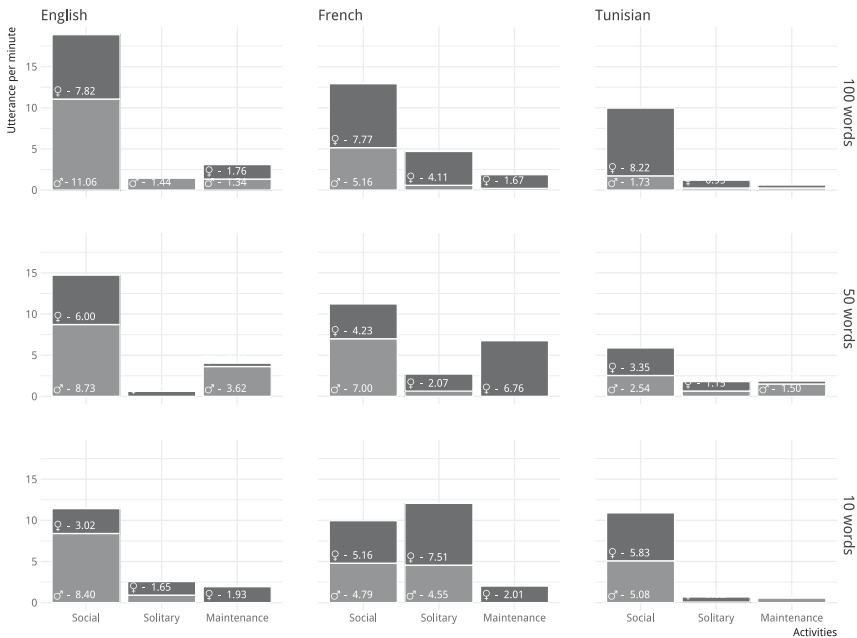


Figure 3: Utterances in child-directed speech per minute as a function of context and language.

3.2.2 Word tokens

While observed proportions of word tokens in CDS (Figure 4) are not considerably different from the above proportions of utterances, they provide evidence for one trend which was less clear when looking at utterances only: the overall amount of speech addressed to girls is greater than that directed to boys. This seems to be the case across languages, and more clearly so in the 100-word level.

3.2.3 Lexical diversity in child-directed speech (VOCD)

Although the D measure is less sensitive to sample size than type-token ratio and therefore more reliable with our data, the measure can only be computed if sufficient data have been gathered (Silverman and Ratner 2002): fifty words are needed for the VOCD command to work in CLAN. Thus, we could not get a reliable D measure for each activity type and child at each level, and we had to pool all three linguistic levels together, as shown in Figure 5.

Caretaker VOCD measures are higher in social contexts than in the other two contexts for American and Tunisian children. For French children however, caretaker VOCD is the highest in the maintenance context. We did not have enough data in solitary context, for the French girl, to compute VOCD. In the social contexts VOCD reaches its highest level for the Tunisian boy and its lowest one for the French boy. Differences across children in maintenance contexts are also quite striking, with measures for the French girl amounting to almost four times those for the

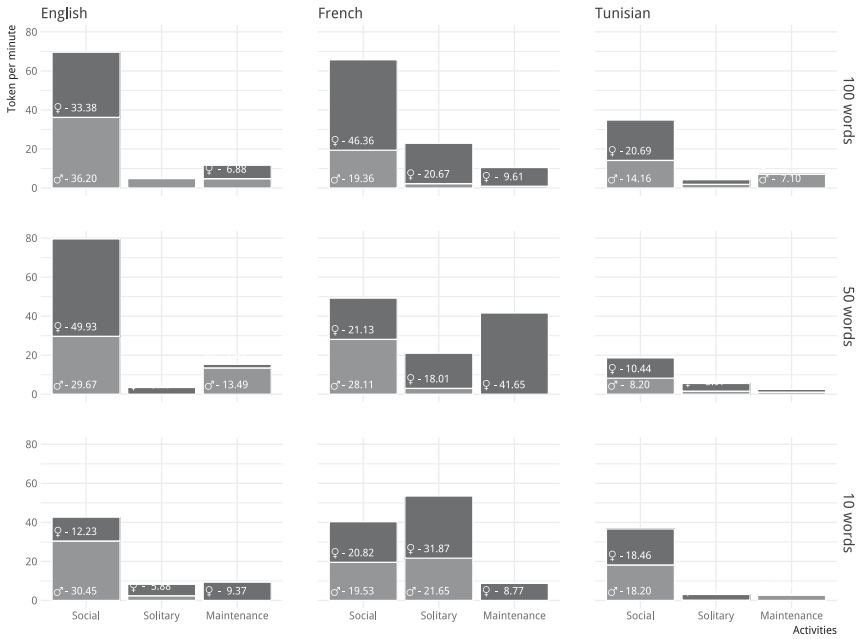


Figure 4: Word tokens in child-directed speech per minute as a function of context and language.

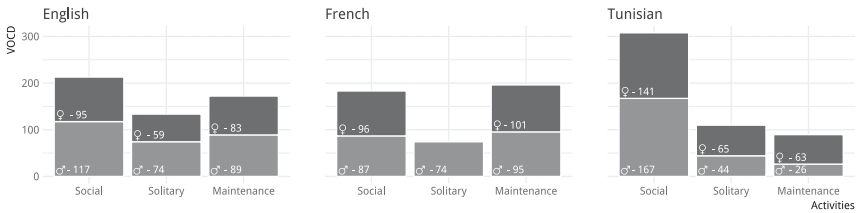


Figure 5: Lexical diversity (VOCD) in child-directed speech as a function of context and language (labels corresponding to different D values).

speech directed to the Tunisian boy, who has the most reduced VOCD in this context. For solitary contexts, we find again the lowest measure with the Tunisian boy, but with much less marked differences.

3.3 Mean length of child-directed utterances

Variability across families and contexts is also observed for mean length of utterances (Figure 6). Overall, no clear developmental progression is observed, and utterances are almost always longer in the maintenance contexts than in the social contexts. Solitary play comes third, except for the French boy and the Tunisian girl. The mean length of utterances directed to the French boy is the same across maintenance

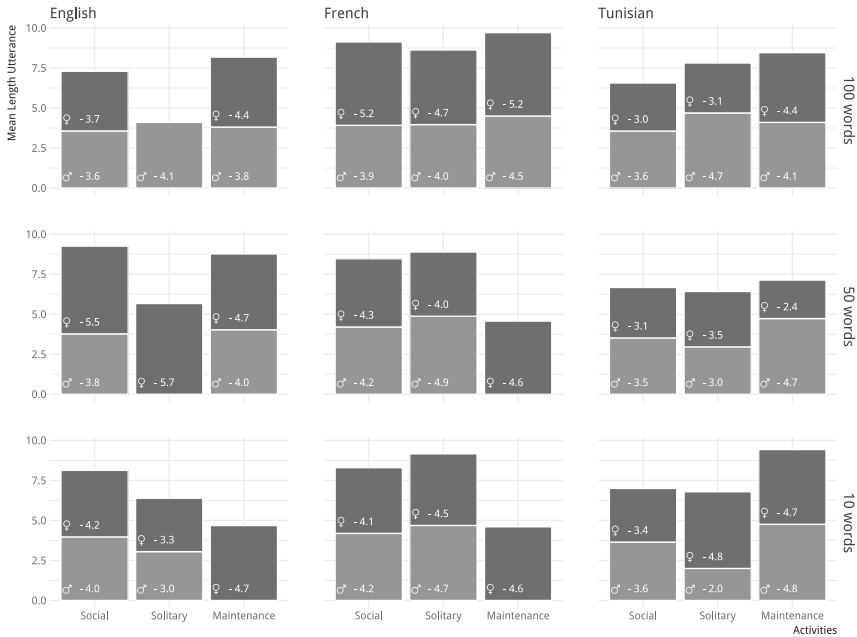


Figure 6: MLU in child-directed speech as a function of context and language (label corresponding to different MLU).

and solitary context (4.8) and longer than in social contexts (4.3). Although observed differences are relatively small, some children stand out. For all three contexts, the French girl's child-directed speech is composed of longer utterances (ranging between 4 and 5.2). On the other hand, the lowest mean length of utterances is found in Tunisian children: in social and maintenance contexts, the Tunisian girl has the shortest utterances directed to her, and for solitary contexts it is the Tunisian boy who receives the shortest utterances.

3.4 Noun-verb proportion in child-directed speech depending on activity

While no strong cross-linguistic differences emerged in previously discussed measures, Tunisian Arabic clearly stands out when looking at proportions of nouns and verbs. More similarities are observed in French and English, with relatively stable ratios and a slightly higher proportion of nouns in all coded data (Figure 7). Although further analyses are needed on Tunisian Arabic, the higher ratios of verbs that are observed in mothers' utterances are reminiscent of Choi's results with Korean data in toy play contexts (Choi 2000: 80) or Tardif's results with Mandarin (Tardif 1996).

In terms of activity types, more variability is observed in maintenance activities. This trend is particularly difficult to analyse since these activities are seldom described in the literature, and notably absent from studies on the acquisition of

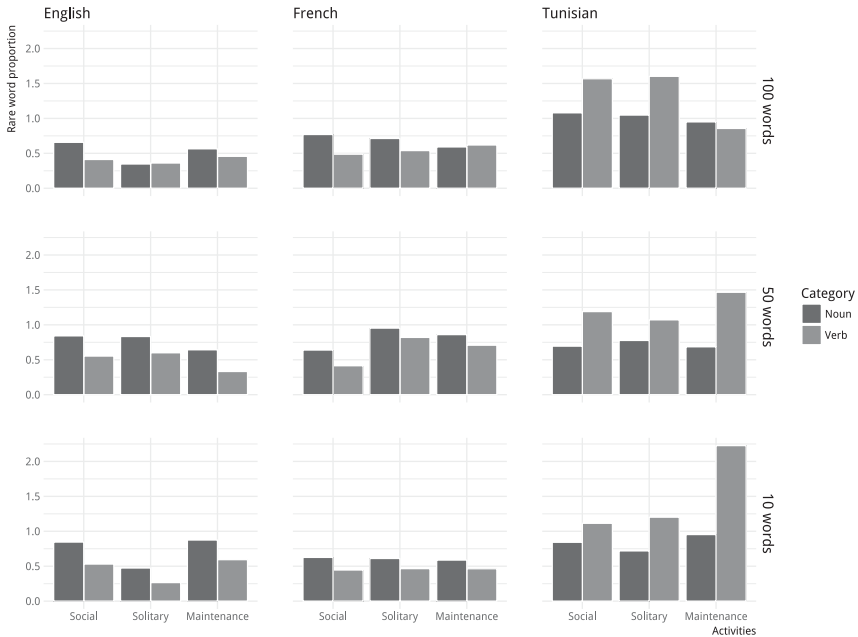


Figure 7: Nouns and verbs (tokens per utterance) in CDS as a function of context, language and developmental level.

nouns and verbs. While the greater variability may stem from the smaller quantity of data, it certainly calls for further investigation.

In terms of developmental progression, the final level stands out across languages and exhibits smaller differences between noun and verb ratios. This is especially true for maintenance activities, again calling for analyses on a bigger sample.

4. DISCUSSION

On the whole, this study of activity contexts has revealed differences in activity duration and distribution across levels and languages, and differences in child-directed speech that are greater across activity contexts than across levels and languages. We did not, however, seek to arrive at broad generalisations. In this section, we look at means, in order to consolidate our results while avoiding overgeneralisations.

4.1 Activity patterns across sessions

The first objective of our study was to convey a detailed picture of activity contexts in longitudinal corpora across languages, and in doing this we observed variety in and across contexts, across children and languages. Table 4 gives an overview of activity patterns by providing mean durations for each child, indicating, language, birth order, and activity.

| Language | Birth order | Activity | Mean Duration |
|------------------|-------------|-------------|---------------|
| American English | 2 | Social | 34.77 |
| | | Solitary | 10.82 |
| | | Maintenance | 7.68 |
| American English | 4 | Social | 38.04 |
| | | Solitary | 7.05 |
| | | Maintenance | 10.62 |
| French | 1 | Social | 16.99 |
| | | Solitary | 20.17 |
| | | Maintenance | 11.76 |
| French | 2 | Social | 22.02 |
| | | Solitary | 24.06 |
| | | Maintenance | 0.67 |
| Tunisian Arabic | 2 | Social | 30.68 |
| | | Solitary | 6.80 |
| | | Maintenance | 1.51 |
| Tunisian Arabic | 1 | Social | 36.35 |
| | | Solitary | 1.69 |
| | | Maintenance | 2.49 |

Table 4: Mean activity duration by language, birth order, and activity.

While means confirm overall trends such as the prevalence of social activities, they also point to notable exceptions. The prevalence of solitary activities in both the French girl's and the French boy's recordings is striking. It could hardly result from linguistic and cultural differences, considering the relative homogeneity of our data and recording procedures. We hypothesise that such differences point to an intricate network of uncontrolled parameters, which characterise naturalistic recordings.

Indeed, the activity contexts in our video recordings depend on various factors like the family schedule, the presence of brothers and sisters, fathers, and other family members, the child's age and the time of recording. Attention must be paid to this complexity if we are to understand the variety of activity contexts in our recordings, especially since all these factors are interconnected. For instance, while birth order may have been used to account for the presence of more solitary activities in first-born children, it is not necessarily the case that siblings are present in the recordings. Besides, variations in the observer's presence and involvement are also likely to have an impact on the diversity of recorded activities. In the French dataset, it was observed that maintenance activities such as diaper change or bath could be recorded only when the observer was there to move the camera around and adjust to what was going on. In most of the American recordings, the observer was there at the beginning of the session but then left the camera on a tripod for an hour, which constrains maintenance activities to snacking when recording in the kitchen or living room – as was often the case. In addition to the constraints listed above, recorded activities seem to depend on the moment of recording (time of

day), which in turn depends on the availability and schedule of each family, as a result of an array of individual and social factors.

The presence of other participants has an influence on interaction, which is why in most studies the dyadic situation is preferred (Leyendecker et al. 1997b). However, this specific situation is not the most frequent for children, and even in recordings such as ours that are primarily focused on dyads, we do not have only dyadic interactions. There are moments when the child is alone, other moments the child is with both parents or even with a sibling. In order to maximise comparability across datasets, we restricted our coding to dyads (child – parents), as is the case in most studies of child-directed speech. However, it is worth noting that as a result of this choice we missed an important part of the children's daily interactional experience (Leyendecker et al. 1997b).

The father's presence is also variable in our data, and when fathers are present they may or may not interact with the child. In our French data, the father of the French boy is present just once among the three sessions, and he doesn't really interact with his child, while the father of the French girl is present for two sessions out of three, and interacts a lot with his daughter. Fathers are not often present and do not often interact with the Tunisian girl and boy: he is seen either as he arrives or as he leaves. In the US data, the girl's father is present in one of our recordings but only because he is having a short conversation with his wife before leaving and going to work, so that there are very few utterances directed to the child. The American boy's father also appears at some points in the recordings but not in any of our analysed files. As for siblings, they are not often present. We have one session in which the French boy's sister was present. In this session we observe social play with sister while mother was busy doing other things, and when joint activity takes place, as in social, toy play and book reading, we have more tantrums and the sister who constantly helps the child and leads the way.

The moment of the day when recording was done is indeed variable, notably because some mothers did not have a full time job: both French mothers were the only caretakers of their children (no nurse or childcare), as was also the case for the two American corpora, where lots of recordings were made in the morning. For Tunisian data, both children were video-recorded by an observer who had a job, so the time of recording was limited to the evening. In addition, French recording is variable for the French girl, mostly in morning or afternoon, but more homogeneous for the French boy where video recording took place in the afternoon. Those differences have a clear impact, notably on maintenance activities: depending on the time of recording, we may have more or less chance to observe bath and mealtime.

Granularity also accounts for our results. Gender-related differences are more likely to appear in fine-grained analyses of subtypes of activities; that is, there is a lot of toy play in the recorded interactions with our Tunisian boy and much more conversation with the Tunisian girl. Contexts such as play, feeding, and caretaking have been shown to have an influence on the quantity of dyadic interaction observed (O'Brien and Nagle 1987, O'Brien et al. 1989) and can mask certain differences which may be found across social classes as well as regarding the extent and nature of maternal discourse (Snow 1972, Wootton 1974, Dunn et al. 1977, Hoff-Ginsberg 1991). These elements are of interest to our results, since we found

considerable variation in maintenance activities across sessions. However, we will need more data to get robust results on linguistic measures in maintenance contexts.

Given the potential influence of social and functional contexts, as described above, we would like to emphasise that selecting and controlling observation procedures in order to gather representative and reliable data remains a challenge, which may be exacerbated by differences in SES³. Consequently, our study reinforces the claim that activity context should be taken into account and could even be used as a means to have valid, reliable and significant observations for estimated child experience (Leyendecker et al. 1997b).

4.2 Child-directed speech description

A second objective of the present study was to grasp fine-grained characteristics of child-directed speech, by comparing activity contexts rather than full recordings. In doing so, we tried to shed further light on the complexity of children's input (see Gentner and Boroditsky 2001) and to single out possible factors of variation.

First, this study has shown differences in total number of utterances addressed to children, as a result of variation in the children's interlocutors. Indeed, the French boy received fewer utterances than the French girl or either of the American children, as is clear from Table 5. This is linked with the fact that the boy's mother regularly talked to the observer, the grandmother when she was present, and she talked a lot on the phone. As a result, less of her time was spent speaking to the French boy. Tunisian children received less than half the number of utterances addressed to all other children. However in the recordings the children are often found interacting with other members of the family, like the grandmother, brothers and sisters. Our failure to take these interactions into account is one clear limitation of our methodology, which could not be overcome because to date we have only partial transcriptions of the data for Tunisian Arabic. Integrating the whole of child-directed speech seems highly desirable in future studies, in order to better grasp who speaks to the children and how, and gain a better picture of activity contexts.

As regards MLU, the tendency for utterances to be longer in maintenance contexts is confirmed by mean values given in Table 6. This is rather unexpected, as MLU is often understood as an index of complexity and richness (Brown 1973, Rondal et al. 1987, Bates and Carnevale 1993).

In order to account for this, we looked at the mothers' productions in maintenance contexts and found that in many situations (including hygiene and snacking) there was a stronger need to provide the child with arguments, and thus use more complex utterances. Table 7 provides two examples from our US data.

Another interesting feature of MLU is that it seems to vary according to the child's gender (Table 8). On the whole, utterances directed to boys are shorter across all contexts, with smaller differences in maintenance contexts, and larger differences in solitary contexts. This may be linked with differences in mothers' verbal

³Our data were homogeneous in this respect, but it remains to be seen whether major differences emerge in activity contexts recorded in longitudinal corpora as a result of SES.

| Language | Gender | Activity | Mean |
|------------------|--------|-------------|------|
| American English | F | Social | 5.61 |
| | | Solitary | 1.13 |
| | | Maintenance | 1.37 |
| American English | M | Social | 9.40 |
| | | Solitary | 1.17 |
| | | Maintenance | 2.48 |
| French | F | Social | 5.72 |
| | | Solitary | 4.57 |
| | | Maintenance | 3.48 |
| French | M | Social | 5.65 |
| | | Solitary | 1.92 |
| | | Maintenance | 0.21 |
| Tunisian Arabic | F | Social | 5.80 |
| | | Solitary | 0.90 |
| | | Maintenance | 0.27 |

Table 5: Mean number of utterances per minute by language, gender, and activity.

| Language | Gender | Activity | Mean |
|------------------|--------|-------------|------|
| American English | F | Social | 4.45 |
| | | Solitary | 4.50 |
| | | Maintenance | 4.59 |
| American English | M | Social | 3.77 |
| | | Solitary | 3.58 |
| | | Maintenance | 3.92 |
| French | F | Social | 4.52 |
| | | Solitary | 4.38 |
| | | Maintenance | 4.79 |
| French | M | Social | 4.10 |
| | | Solitary | 4.51 |
| | | Maintenance | 4.50 |
| Tunisian Arabic | F | Social | 3.16 |
| | | Solitary | 3.79 |
| | | Maintenance | 3.81 |
| Tunisian Arabic | M | Social | 3.58 |
| | | Solitary | 3.22 |
| | | Maintenance | 4.53 |

Table 6: Mean Length Utterance by language, gender, and activity.

style (responsiveness and directiveness) which have been evidenced as a function of gender: mothers' responsiveness was shown to be stronger with girls while directiveness prevailed with boys (Flynn and Masur 2007).

| Maintenance (snacking) with English girl, 100-word level: | |
|--|--|
| Mother: | I'll toast you some little bagel Okay Oh xxx they're yucky Can't toast you the little bagels (be)cause they're all moldy Let's see |
| Maintenance (dinner) with US boy 50-word level | |
| Mother: | Wait a minute; we have some soup okay, next bite. Thank you, that's delicious. Thank you for sharing, have some soup now. |
| Boy: | No! |
| Mother: | No soup, oh. You have to have some soup. How will you get muscles? Where's your muscles? I don't feel any muscles. |

Table 7: Examples of utterances during maintenance activity.

| Gender | Activity | Mean |
|--------|-------------|------|
| F | Social | 4.05 |
| | Solitary | 4.19 |
| | Maintenance | 4.40 |
| M | Social | 3.82 |
| | Solitary | 3.79 |
| | Maintenance | 4.32 |

Table 8: Mean Length of Utterance by gender and activity.

Finally, we used nouns and verbs as a well-described index of cross-linguistic differences, and one that has often been shown to vary across contexts. We expected to find more nouns in the speech of American caretakers than in that of speakers from the other two languages, and indeed this was the case for most of our data. Overall, if previous studies have shown that in several linguistic communities mothers used more nouns than verbs (see Fernald and Morikawa 1993, Goldfield 1993, Poulin-Dubois et al. 1995, Gopnik et al. 1996, Tardif et al. 1997, Choi 1998, Kim et al. 2000) the reverse seems to be true for Tunisian Arabic, which to our knowledge has never been studied in terms of noun and verb proportions in CDS. However, contrary to recent studies about nouns and verbs which situate linguistics measures in context (Choi 2000, Altunkamiş et al. 2014) we find more homogeneity across activities than across languages. Our results are not fully comparable, however, since we are dealing with longitudinal recordings that were not structured according to activity

types. We also collapsed toy play and book reading activities in the social context. With more data, it will be possible to look at subtypes, which may give more illuminating results for nouns and verbs.

4.3 Rare words

Before concluding, we would like to discuss more qualitative analyses which were conducted in order to shed light on lexical diversity. Because lexical diversity measures did not yield the expected results, we looked for rare words in CDS as an indication of increased diversity (Beals 1997), and tried to see if their use was tied to specific activity contexts. To date, rare words have been studied with varied objectives and methods: while Snow and Beals (2006) focused on family dinners and started from an existing list, Parisse (2014) started from dense corpora to extract word frequencies and isolated rare words within a zipfian distribution. In order to find rare words in our data, we used a method that could be replicated cross-linguistically and did not depend on pre-existing lists. We computed the frequencies of the words occurring in our data and compared them to frequencies across all 14 transcripts available for each dataset (i.e., our French, English and Tunisian transcriptions).⁴ Table 9 gives examples of rare words found in our French data.

In order to determine which context our rare words predominantly appeared in, we computed means according to language and activity (Table 10). Different means appear for Tunisian due to the lack of transcribed data⁵ in which to find rare words. In the other two languages, however, regularities emerge, with a greater proportion of rare words found in English social activities and French solitary activities.

Looking at the detailed occurrences at each level (see Table 11), we see more occurrences of rare words in maintenance contexts at the 10-word level for English children, and in solitary contexts again at the 10-word level for French children.

Besides, the noun bias is confirmed in both our French and English rare word lists. Only our Tunisian list contains more verbs than nouns, in accordance with the higher verb ratios found. When we add all developmental levels per children, French and American caretakers' rare words contain more nouns than verbs, but the opposite is true for Tunisian caretakers.

Finally, in terms of the pragmatic value of utterances, it is worth emphasising that in our findings, rare words appear most often when the mother comments on the child activity. Using existing pragmatic coding in other datasets might enable us to see whether it is indeed the case that mothers' well-described vocabulary simplification (Hayes and Ahrens 1988) stops in these small pieces of speech where they do not address the child directly.

⁴With the exception of Tunisian Arabic, where only three fully transcribed sessions were available.

⁵Thus, we could not properly identify rare words in our Tunisian Arabic dataset, and ended up with lists that are considerably longer than in the other two languages, and less tied to specific contexts.

| | |
|---|------------------------|
| Craspouillou ‘funny nickname meaning dirty little one’ | (solitary context) |
| Mother comments upon child’s activity while child explores environment alone: | |
| Qu’est+ce tu fais craspouillou là à baver ? | |
| ‘What are you doing here drooling around, dirty little one?’ | |
| Se déchaîner ‘to be frantic’ | (solitary context) |
| Child is overexcited and his mother comments upon his activity: | |
| Et mais tu te déchaîne toi ce matin, là hein. | |
| ‘You’re frantic this morning huh.’ | |
| S’extasier ‘to marvel’ | (social play with toy) |
| The tower is rising and he will soon be able to destroy it: | |
| Commence pas à t’extasier là. | |
| ‘Don’t start marveling at it now.’ | |
| Gambader ‘to gambol’ | (bookreading context) |
| Allez Bibi voudrait bien gambader comme ses amis les moutons. | |
| ‘Well, it seems Bibi would like to gambol like his friends the sheep.’ | |
| Sketch ‘comedy sketch’ | (social routine) |
| Child begins to be very excited: | |
| Je te laisse là faire ton ... ton sketch . | |
| ‘I’ll let you do your ...your comedy sketch.’ | |
| Souk ‘mess or chaos’ | (exploration context) |
| Child moves all his toys out, mother comments upon his activity: | |
| Et ben dis+donc t’as mis du bazar toi hein en deux minutes là. C’est le souk. | |
| ‘Oh well, you made a real mess in here, in just two minutes huh. It’s chaos.’ | |
| Sac+à+main ‘hand-bag’ | (solitary play) |
| Mother comments on her child’s action: | |
| T’as pris ton sac+à+main ? | |
| ‘Did you take your handbag?’ | |
| Disparition ‘disappearance’ | (bookreading context) |
| pas plus que dans la cuisine d’ailleurs c’est étrange cette disparition. | |
| ‘not in the kitchen either by the way, what a strange disappearance.’ | |

Table 9: Example of rare words in the French boy’s corpus.

| Language | Activity | Mean |
|------------------|-------------|------|
| American English | Social | 0.07 |
| | Solitary | 0.04 |
| | Maintenance | 0.03 |
| French | Social | 0.02 |
| | Solitary | 0.04 |
| | Maintenance | 0.02 |
| Tunisian Arabic | Social | 0.28 |
| | Solitary | 0.20 |
| | Maintenance | 0.24 |

Table 10: Mean rare word percent age by language and activity.

| Language | Level | Activity | Mean |
|----------|-----------|-------------|------|
| English | 100 words | Social | 0.08 |
| | | Solitary | 0.03 |
| | | Maintenance | 0.04 |
| | 50 words | Social | 0.06 |
| | | Solitary | 0.02 |
| | | Maintenance | 0.02 |
| | 10 words | Social | 0.06 |
| | | Solitary | 0.05 |
| | | Maintenance | 0.05 |
| French | 100 words | Social | 0.04 |
| | | Solitary | 0.04 |
| | | Maintenance | 0.02 |
| | 50 words | Social | 0.01 |
| | | Solitary | 0.03 |
| | | Maintenance | 0.03 |
| | 10 words | Social | 0.02 |
| | | Solitary | 0.05 |
| | | Maintenance | 0.01 |
| Tunisian | 100 words | Social | 0.32 |
| | | Solitary | 0.21 |
| | | Maintenance | 0.30 |
| | 50 words | Social | 0.28 |
| | | Solitary | 0.20 |
| | | Maintenance | 0.27 |
| | 10 words | Social | 0.24 |
| | | Solitary | 0.19 |
| | | Maintenance | 0.14 |

Table 11: Mean rare word percentage by language, level, and activity.

4.4. Conclusion

The main finding of our study is that apparent subcultural differences are maximised by focusing on single contexts and minimised by averaging across a variety of naturally occurring contexts. Our study confirms that observation across a variety of uncontrolled activity contexts raises comparability issues (Stevenson et al. 1986, Lewis and Gregory 1987, O'Brien and Nagle 1987, Lewedag et al. 1994). It suggests that we need more fine-grained observations of language acquisition and input in cross-cultural studies. This is especially true for our category of maintenance, which corresponds to activities that are seldom described in the literature: we have shown that child-directed speech in maintenance activities was rich and varied. Our analyses thus call for more detailed studies of feeding or bath-time activities. This could be done in naturalistic yet standardised situations (see Bornstein and Haynes 1998) and perhaps more specifically by zooming in on mealtime interactions (Pan et al. 1999) to overcome the obstacles described in this study. Ultimately, our study questions the level of granularity that is desirable when studying language development with longitudinal data. Finer categories could indeed be isolated even within our subcategories, since mothers or children sometimes engage in very short activities before returning to what they were doing. In the present study, we proposed a three-fold coding system, which evidenced major trends in longitudinal corpora, where social activities dominate, and solitary and maintenance activities are much less frequent. We hope that this finding, together with the variety we discussed, will bring new insights into existing data and foster data collection in less represented activity contexts.

CORPORA

Human Speechome Corpus <<https://www.media.mit.edu/cogmac/projects/hsp.html>>

Providence corpus <<https://phonbank.talkbank.org/access/Eng-NA/Providence.html>>

PREMS corpus <<http://www.dcl.cnrs.fr/projets/prems/index.asp?Langue=EN&Page=Home>>

OHLL corpus <<https://phonbank.talkbank.org/access/French/KernFrench.html>>

REFERENCES

- Altunkamış, Feyza, Sophie Kern, and Hatice Sofu. 2014. When context matters more than language: Verb or noun in French and Turkish caregiver speech. *First Language* 34(6): 537–550.
- Amayreh, Mousa M. 2003. Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research* 46(3): 517–529.
- Au, Terry Kit-Fong, M. Dapretto, and Y. K. Song. 1994. Input Vs constraints: Early word acquisition in Korean and English. *Journal of Memory and Language* 33(5): 567–582.
- Badry Zalami, Fatima. 2006. First language acquisition. In *Encyclopedia of Arabic Language and Linguistics*, vol 2, 96–104. Boston: Brill.
- Bassano, Dominique. 1998. L'élaboration du lexique précoce chez l'enfant français : structure et variabilité. *Enfance* 51(4): 123–153.

- Bassano, Dominique. 2000. Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language* 27(3): 521–559.
- Bates, Elizabeth, Inge Bretherton, and Lynn Sebestyen Snyder. 1991. *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bates, Elizabeth, and George F. Carnevale. 1993. New directions in research on language development. *Developmental Review* 13(4): 436–470.
- Bates, Elizabeth, Virginia Marchman, Donna Thal, Larry Fenson, Philip Dale, Steven J. Reznick, Judy Reilly, and Jeff Hartung. 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* 21(1): 85–123.
- Beals, Diane E. 1997. Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language* 24(3): 673–694.
- Benedict, Helen. 1979. Early lexical development: Comprehension and production. *Journal of Child Language* 6(2): 183–200.
- Bloom, Lois. 1973. *One word at a time: The use of single-word utterances before syntax*. The Hague: Mouton.
- Bornstein, Marc H., Catherine S. Tamis-LeMonda, Joseph Tal, Pamela Ludemann, Sueko Toda, Charles W. Rahn, Marie-Germaine Pêcheux, Hiroshi Azuma, and Danya Vardi. 1992. Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development* 63(4): 808–821.
- Bornstein, Marc H., and Maurice O. Haynes. 1998. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development* 69(3): 654–671.
- Bornstein, Marc H., Maurice O. Haynes, and Kathleen M. Painter. 1998. Sources of child vocabulary competence: A multivariate model. *Journal of Child Language* 25(2): 367–393.
- Bornstein, Marc H., Linda R. Cote, Sharone Maital, Kathleen Painter, Sung-Yun Park, Liliana Pascual, Marie-Germaine Pêcheux, Josette Ruel, Paola Venuti, and Andre Vyt. 2004. Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development* 75(4): 1115–1139.
- Brown, Roger. 1973. Development of the first language in the human species. *American Psychologist* 28(2): 97–106.
- Bruner, Jerome. 1981. The social context of language acquisition. *Language and Communication* 1(2–3): 155–178.
- Cartmill, Erica A., Benjamin F. Armstrong, Lila R. Gleitman, Susan Goldin-Meadow, Tamara N. Medina, and John C. Trueswell. 2013. Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences* 110: 11278–11283.
- Caselli, Maria Cristina, Elizabeth Bates, Paola Casadio, Judi Fenson, Larry Fenson, Lisa Sanderl, and Judy Weir. 1995. A cross-linguistic study of early lexical development. *Cognitive Development* 10(2): 159–199.
- Caselli, Cristina, Paola Casadio, and Elizabeth Bates. 1999. A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* 26(1): 69–111.
- Choi, Soonja. 1998. Verbs in early lexical and syntactic development in Korean. *Linguistics* 36(4): 755–780.
- Choi, Soonja. 2000. Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language* 27(1): 69–96.
- Choi, Soonja, and Alison Gopnik. 1995. Early acquisition of verbs in Korean: A cross-linguistics study. *Journal of Child Language* 22(3): 497–529.

- Clark, Eve V. 2009. *First language acquisition*. Cambridge: Cambridge University Press.
- Demuth, Katherine, Jennifer Culbertson, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech* 49 (2): 137–173.
- Dunn, Judy, Carol Wooding, and Joan Herman. 1977. Mothers' speech to young children: Variation in context. *Developmental Medicine and Child Neurology* 19(5): 629–638.
- Ely, Richard, Jean Berko Gleason, Ann MacGibbon, and Elena Zaretsky. 2001. Attention to language: Lessons learned at the dinner table. *Social Development* 10(3): 355–373.
- Eriksson, Mårten, Peter B. Marschik, Tiia Tulviste, Margareta Almgren, Miguel Pérez Pereira, Sonja Wehberg, Ljubica Marjanovič-Umek, Frédérique Gayraud, Melita Kovacevic, and Carlos Gallego. 2012. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology* 30(2): 326–343.
- Farran, Dale C., and Ron Haskins. 1980. Reciprocal influence in the social interactions of mothers and three-year-old children from different socioeconomic backgrounds. *Child Development* 51(3): 780–791.
- Fenson, Larry, Philip S. Dale, Steven J. Resnick, Donna Thal, Elizabeth Bates, Jeffrey P. Hartung, Stephen Pethick, and Judy S. Reilly. 1993. *MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.
- Fenson, Larry, Philip S. Dale, Steven J. Reznick, Elizabeth Bates, Donna J. Thal, Stephen J. Pethick, Michael Tomasello, Carolyn B. Mervis, and Joan Stiles. 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development* 59: 185.
- Fernald, Anne, and Hiromi Morikawa. 1993. Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development* 64(3): 637–656.
- Florin, Agnès. 1999. *Le développement du langage*. Paris: Dunod DL.
- Flynn, Valerie, and Elise Frank Masur. 2007. Characteristics of maternal verbal style: Responsiveness and directiveness in two natural contexts. *Journal of Child Language* 34(3): 519–543.
- Gentner, Dedre. 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Center for the Study of Reading Technical Report 257, University of Illinois. Available at <<http://hdl.handle.net/2142/17514>>.
- Gentner, Dedre, and Lera Boroditsky. 2001. Individuation, relativity, and early word learning. In *Language acquisition and conceptual development*, ed. Melissa Bowerman and Stephen Levinson, 215–256. Cambridge: Cambridge University Press.
- Goldfield, Beverly A. 1993. Noun bias in maternal speech to one-year-olds. *Journal of Child Language* 20(1): 85–99.
- Goldfield, Beverly A. 2000. Nouns before verbs in comprehension vs. *production: The view from pragmatics*. *Journal of Child Language* 27(3): 501–520.
- Golinkoff Michnick, Roberta, Dilara Deniz Can, Melanie Soderstrom, and Kathy Hirsh-Pasek. 2015. (Baby)talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science* 24(5): 339–344.
- Goodman, Judith C., Philip S. Dale, and Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35(3): 515–531.
- Gopnik, Alison, and Soonja Choi. 1990. Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language* 10(30): 199–215.

- Gopnik, Alison, Soonja Choi, and Therese Baumberger. 1996. Cross-linguistic differences in early semantic and cognitive development. *Cognitive Development* 11(2): 197–225.
- Haden, Catherine A., and Robyn Fivush. 1996. Contextual variation in maternal conversational styles. *Merrill-Palmer Quarterly* 42(2): 200–227.
- Hamilton, Antonia, Kim Plunkett, and Graham Schafer. 2000. Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language* 27(3): 689–705.
- Hart, Betty, and Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing.
- Hayes, Donald P., and Margaret G. Ahrens. 1988. Vocabulary simplification for children: A special case of ‘motherese’? *Journal of Child Language* 15(2): 395–410.
- Hindman, Annemarie H., Lori E. Skibbe, and Tricia D. Foster. 2014. Exploring the variety of parental talk during shared book reading and its contributions to preschool language and literacy: Evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Reading and Writing* 27(2): 287–313.
- Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26(1): 55–88.
- Hoff, Erika, Brett Laursen, Twila Tardif, and Marc H. Bornstein. 2002. Socioeconomic status and parenting. In *Handbook of parenting, Vol 2: Biology and ecology of parenting* (2nd ed.), 231–252. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hoff-Ginsberg, Erika. 1991. Mother-child conversation in different social classes and communicative settings. *Child Development* 62(4): 782–796.
- Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology* 27(2): 236–248.
- Jackson-Maldonado, Donna, Donna Thal, Virginia Marchman, Elizabeth Bates, and Vera Gutierrez-Clellen. 1993. Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language* 20(3): 523–549.
- Kauschke, Christina, and Christoph Hofmeister. 2002. Early lexical development in German: A study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language* 29(4): 735–757.
- Kelly, Barbara F., William Forshaw, Rachel Nordlinger, and Gillian Wigglesworth. 2015. Linguistic diversity in first language acquisition research: Moving beyond the challenges. *First Language* 35(4–5): 286–304.
- Kern, Sophie. 2005. De l’universalité et des spécificités du développement langagier précoce. In *Aux origines des langues et du langage*, ed. Jean-Marie Hombert, 270–291. Paris: Fayart.
- Kern, Sophie. 2010. Les premiers mots du jeune enfant français: analyse quantitative et qualitative du vocabulaire réceptif et productif des deux premières années de vie. *Rééducation orthophonique* 244: 149–165.
- Kern, Sophie, Florence Chenu, and Fayza Türkay. 2012. Poids du contexte sur la répartition noms/verbes dans le langage adressé à l’enfant en français et en turc. *SHS Web of Conferences* 1: 1607–1622.
- Kern, Sophie, Barbara L. Davis, and Inge Zink. 2009. From babbling to first words in four languages: Common trends, cross language and individual differences. In *Becoming eloquent: Advances in the emergence of language, human cognition, and modern culture*, ed. Francesco d’Errico and Jean-Marie Hombert, 205–332. Amsterdam: Benjamins.

- Kim, Mikyong, Karla McGregor, and Cynthia K. Thompson. 2000. Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. *Journal of Child Language* 27(2): 225–254.
- Kloth, Saskia, Peggy Janssen, Floris Kraaimaat, and Gene J. Brutton. 1998. Communicative styles of mothers interacting with their preschool-age children: A factor-analytic study. *Journal of Child Language* 25(1): 149–168.
- Lewedag, Vanessa L., Kimbrough D. Oller, and Michael P. Lynch. 1994. Infants' vocalization patterns across home and laboratory environments. *First Language* 14(40): 49–65.
- Lewis, Charlie, and Susan Gregory. 1987. Parents' talk to their infants: The importance of context. *First Language* 7(21): 201–216.
- Leyendecker, Birgit, Michael E. Lamb, and Axel Schölmerich. 1997a. Studying mother-infant interaction: The effects of context and length of observation in two subcultural groups. *Infant Behavior and Development* 20(3): 325–337.
- Leyendecker, Birgit, Michael E. Lamb, Axel Schölmerich, and Miranda Del Fricke. 1997b. Contexts as moderators of observed interactions: A study of Costa Rican mothers and infants from differing socioeconomic backgrounds. *International Journal of Behavioral Development* 21(1): 15–34.
- MacWhinney, Brian. 2009. *The CHILDES project: Tools for analyzing talk*. 3. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maital, Sharone L., Esther Dromi, Abraham Sagi, and Marc H. Bornstein. 2000. The Hebrew Communicative Development Inventory: Language-specific properties and cross-linguistic generalizations. *Journal of Child Language* 27(1): 43–67.
- Malvern, David, Brian Richards, Ngoni Chipere, and Pilar Durán. 2004. *Lexical diversity and language development*. New York: Palgrave Macmillan.
- Masur, Elise F., and Jean Berko Gleason. 1980. Parent-child interaction and the acquisition of lexical information during play. *Developmental Psychology* 16(5): 404–409.
- Newland, Lisa A., Lori A. Roggman, and Lisa K. Boyce. 2001. The development of social toy play and language in infancy. *Infant Behavior and Development* 24(1): 1–25.
- O'Brien, Marion, Jeanne M. Johnson, and Diana Anderson-Goetz. 1989. Evaluating quality in mother-infant interaction: Situational effects. *Infant Behavior and Development* 12(4): 451–464.
- O'Brien, Marion, and Keith J. Nagle. 1987. Parents' speech to toddlers: The effect of play context. *Journal of Child Language* 14(2): 269–279.
- Omar, Margaret K. 2007. *The acquisition of Egyptian Arabic as a native language*. Washington: Georgetown University Press.
- Pan, Barbara Alexander, Rivka Y. Perlmann, and Catherine E. Snow. 1999. Food for thought: Dinner table as a context for observing parent-child discourse. In *Methods for studying language production*, ed. Lise Menn and Nan Bernstein Ratner, 205–222. Oxford: Psychology Press.
- Parisse, Christophe. 2014. Événements langagiers rares et acquisition du langage. In *Congrès mondial de linguistique française*, <<https://doi.org/10.1051/shsconf/20140801339>>.
- Parisse, Christophe, and Marie Thérèse Le Normand. 2000. How children build their morpho-syntax: The case of French. *Journal of Child Language* 27(2): 267–292.
- Pellegrini, Anthony D., Franck Symons, and John Hoch. 2004. *Observing children in their natural worlds: A methodological primer*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Poulin-Dubois, Diane, Susan Graham, and Lorrie Sippola. 1995. Early lexical development: The contribution of parental labelling and infants' categorization abilities. *Journal of Child Language* 22(2): 325–343.

- Raikes, Helen, Barbara Alexander, Gayle Luze, Catherine S. Tamis-LeMonda, Jeanne Brooks-Gunn, Jill Constantine, Louisa Banks Tarullo, Abigail Raikes, and Eileen T. Rodriguez. 2006. Mother-child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development* 77(4): 924–953.
- Reese, Elaine, and Adell Cox. 1999. Quality of adult book reading affects children's emergent literacy. *Developmental Psychology* 35(1): 20–28.
- Rondal, Jean A., Martine Ghiotto, Serge Bredart, and Jean-François Bachelet. 1987. Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language* 14(3): 433–446.
- Rose, Yvan, Brian MacWhinney, Rodrigue Byrne, Gregory Hedlund, Keith Maddocks, Philip O'Brien, and Todd Wareham. 2005. Introducing Phon: A software solution for the study of phonological acquisition. In *Proceedings of the 30th Annual Boston University Conference on Language Development*, ed. David Bamman, Tatiana Magnitskaia, and Colleen Zaller, 489–500. Boston: Cascadilla Press.
- Rowe, Meredith L. 2008. Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language* 35(1): 185–205.
- Roy, Brandon C. 2013. The birth of a word. Doctoral dissertation, Massachusetts Institute of Technology.
- Roy, Brandon C. 2014. Précis of The birth of a word (summary of a TED talk). <http://web.media.mit.edu/~bcroy/papers/bcroy-phd_precis.pdf>.
- Roy, Brandon C., Michael C. Frank, and Deb Roy. 2009. Exploring word learning in a high-density longitudinal corpus. *Proceedings of the Annual Meeting of the Cognitive Science Society* 31. Amsterdam: Cognitive Science Society.
- Roy, Brandon C., Michael C. Frank, and Deb Roy. 2012. Relating activity contexts to early word learning in dense longitudinal data. Proceedings of the 34th Annual Meeting of the Cognitive Science Society. Available at <http://socialmachines.org/papers/cogsci2012_RFR_fix.pdf>
- Roy, Brandon C., Michael C. Frank, Philip DeCamp, Matthew Miller, and Deb Roy. 2015. Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences* 112: 12663–12668.
- Roy, Deb. 2009. New horizons in the study of child language acquisition. Proceedings of Interspeech 2009, Brighton, UK: <<http://dspace.mit.edu/handle/1721.1/65900>>.
- Roy, Deb, Rupal Patel, Philip DeCamp, Rony Kubat, Michael Fleischman, Brandon C. Roy, Nikolaos Mavridis, Stefanie Tellex, Alexia Salata, Jethran Guinness, Michael Levit, and Peter Gornia. 2006a. The Human Speechome Project. In *Symbol Grounding and Beyond*, ed. Paul Vogt, Yuuya Sugita, Elio Tuci and Chrystopher Nehaniv, 192–196. Berlin: Springer.
- Roy, Deb, Rupal Patel, Philip DeCamp, Rony Kubat, Michael Fleischman, Brandon C. Roy, Nikolaos Mavridis, Stefanie Tellex, Alexia Salata, Jethran Guinness, Michael Levit, and Peter Gornia. 2006b. The Human Speechome Project. Paper presented at the 28th annual Cognitive Science Conference, Vancouver.
- Silverman, Stacy, and Nan Bernstein Ratner. 2002. Measuring lexical diversity in children who stutter: Application of VOCD. *Journal of Fluency Disorders* 27(4): 289–304.
- Snow, Catherine E. 1972. Mothers' speech to children learning language. *Child Development* 43(2): 549–565.
- Snow, Catherine E., A. Arlman-Rupp, Y. Hassing, J. Jobse, J. Joosten, and J. Vorster. 1976. Mothers' speech in three social classes. *Journal of Psycholinguistic Research* 5(1): 1–20.
- Snow, Catherine E., and Diane E. Beals. 2006. Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development* 111: 51–66.

- Snow, Catherine E., Clara Dubber, and Akke De Blauw. 1982. Routines in mother-child interaction. in *The language of children reared in poverty: Implications for evaluation and intervention*, ed. Lynne Vernon-Feagans and Dale Clark Farrar, 53–72. New York: Academic Press.
- Stevenson, Marguerite B., Lewis Leavitt, Mary A. Roach, Robin S. Chapman, and Jon F. Miller. 1986. Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research* 15(5): 451–461.
- Tamis-LeMonda, Catherine S., Marc H. Bornstein, Lisa Baumwell, and Amy Melstein Damast. 1996. Responsive parenting in the second year: Specific influences on children's language and play. *Early Development and Parenting* 5(4): 173–183.
- Tamis-LeMonda, Catherine S., March H. Bornstein, Lisa Cyphers, Sueko Toda, and Misako Ogino. 1992. Language and play at one year: A comparison of toddlers and mothers in the United States and Japan. *International Journal of Behavioural Development* 15(1): 19–42.
- Tamis-LeMonda, Catherine S., Jacqueline D. Shannon, Natasha J. Cabrera, and Michael E. Lamb. 2004. Fathers and mothers at play with their 2- and 3-year-olds: Contributions to language and cognitive development. *Child Development* 75(6): 1806–1820.
- Tardif, Twila. 1996. Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* 32(3): 492–504.
- Tardif, Twila, Paul Fletcher, Weilan Liang, Zhixiang Zhang, Niko Kaciroti, and Virginia A. Marchman. 2008. Baby's first 10 words. *Developmental Psychology* 44(4): 929–938.
- Tardif, Twila, Susan A. Gelman, and Fan Xu. 1999. Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development* 70(3): 620–635.
- Tardif, Twila, Marilyn Shatz, and Letitia Naigles. 1997. Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language* 24(3): 535–565.
- Thordardottir, Elin T., and Susan Ellis Weismer. 1996. Language assessment via parent report: Development of a screening instrument for Icelandic children. *First Language* 16(48): 265–285.
- Vanormelingen, Liesbeth, and Steven Gillis. 2016. The influence of socio-economic status on mothers' volubility and responsiveness in a monolingual Dutch-speaking sample. *First Language* 36(2): 140–156.
- Vihman, Marilyn May, and Lorraine McCune. 1994. When is a word a word? *Journal of Child Language* 21(3): 517–542.
- Weisleder, Adriana, and Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science* 24 (11): 2143–2152.
- Weizman, Zehava Oz, and Catherine E. Snow. 2001. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology* 37(2): 265–279.
- Wiley, Angela R., Cecilia Shore, and W. E. Dixon. 1989. Situational differences in the type of utterances mothers use with thirteen-month-old children. Presented at the Society for Research in Child Development, Kansas City, MO. <<http://eric.ed.gov/?id=ED309854>>.
- Wootton, Antony J. 1974. Talk in the homes of young children. *Sociology* 8(2): 277–295.
- Yont, Kristine M., Catherine E. Snow, and Lynne Vernon-Feagans. 2003. The role of context in mother-child interactions: An analysis of communicative intents expressed during toy play and book reading with 12-month-olds. *Journal of Pragmatics* 35(3): 435–454.

APPENDIX 1: EXAMPLES OF ACTIVITIES IN FRENCH CORPUS (FRENCH GIRL AT 100-WORD LEVEL)

Exploratory activity (exploration)

- 10 Mother [qu'est+ce+que t' as pris dans ta main ?] 'What have you got in your hand?'
 11 Mother [t' as des fleurs ?] 'You have flowers?'
 12 Mother [oh elles sont belles !] 'Oh it's beautiful!'
 13 Mother [on va les mettre dans la poubelle les fleurs ?] 'We will put them in the trash, the flowers?'

Exploratory activity (solitary play)

- 248 Mother [on on dessine !] 'We we're drawing!'
 249 Mother [tu t' appliques ?] 'You apply yourself?'
 250 Mother [c'est quoi ce que tu dessines ?] 'What is it that you're drawing?'
 251 Mother [c'est ton papa ?] 'Is it your dad?'

Social activity (manners routine)

- 21 Child [yyy.] 'yyy.'
 22 Mother [tu nous dis bonjour ?] 'You say hello?'
 23 Mother [tu fais un petit coucou ?] 'You do a little hello?'
 24 Mother [non ?] 'No?'

Social activity (teach routine)

- 68 Mother [combien il+y+en+avait de fleur PRENOM ?] 'How many flowers were there NAME?'
 69 Child [a plus.] 'Is no more.'
 70 Mother [il+y+en+a plus ?] 'There is no more?'
 71 Mother [oui non il+y+en+a plus maintenant.] 'Yes no now there is no more.'

Maintenance activity (toilet-training)

- 472 Mother [tu veux le pot ?] 'You want the potty?'
 473 Child [non.] 'No.'
 474 Mother [maman elle va te chercher le pot ?] 'Mom she will get you the potty?'

Maintenance activity (bath routine)

- 1054 Mother [allez maman elle te met dans l' eau ?] 'Go mom she'll put you in the water?'
 1055 Child [mh.] 'Hum.'
 1056 Mother [on grimpe ?] 'We're climbing?'
 1057 Mother [hop+là.] 'Up here.'
 1058 Mother [tiens voilà.] 'Keep there.'
 1059 Mother [voilà mademoiselle PRENOM dans le bain .] 'Here we go miss NAME, in the bath.'