

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

Ludivine Glas, Caroline Rossi, Rim Hamdi-Sultan, Cédric Batailler, Hacène Bellemmouche

▶ **To cite this version:**

Ludivine Glas, Caroline Rossi, Rim Hamdi-Sultan, Cédric Batailler, Hacène Bellemmouche. Activity types and child-directed speech: a comparison between French, Tunisian Arabic and English.. Canadian Journal of Linguistics / Revue canadienne de linguistique, 2018. halshs-01740841

HAL Id: halshs-01740841
https://shs.hal.science/halshs-01740841

Submitted on 5 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 ILCEA4 (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 LPNC (UMR5105, Université Grenoble-Alpes)
cedric.batailler@univ-grenoble-alpes.fr
HACENE BELLEMMOUCHE
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 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 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, cross-linguistic 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 révèlera-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

1.1. Language development: Similarity and variability across children and languages

Cross-linguistic comparisons have evidenced similarities across children's expressive vocabularies (e.g. Caselli et al. 1995, 1999, Gopnik and Choi 1990, Hamilton et al. 2000, Jackson-Maldonado et al. 1993, Maital et al. 2000, Thordardottir and Weismer 1996). In fact, as they acquire their mother tongue, all children go through the same developmental steps: vocabulary size and age of acquisition have been shown to be very similar across for, French and English (Poulin-Dubois et al. 1995), but also in other Romance languages such as Mexican Spanish (Jackson-Maldonado et al. 1993), Italian (Caselli et al. 1995). Two different steps are generally acknowledged across languages. These steps are characterised by different rhythms in new word acquisition and by the use of words belonging to different grammatical categories. Indeed, despite cross-linguistic or inter-individual differences, on the whole the same reorganisations in children's vocabulary occur with regard to word class between the age 12 and 36 month. A first step is characterized by rather slow acquisition of words that are mostly nouns (Bassano 1998, Bornstein et al. 2004, Caselli et al. 1995, Jackson-Maldonado et al. 1993, Kauschke and Hofmeister 2002, Maital et al. 2000), followed by much quicker lexical acquisition spurt and first word combinations.

Nevertheless, variability among children has been evidenced as a result various factors, including gender (girls having slightly larger productive vocabularies, as shown e.g. by Eriksson et al. 2012 using adapted CDIs for 10 non-English language communities) birth rank (see 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 SES, the language being acquired and the interactions or activity types. The impact of endogenous factors as well as SES has been

controlled in our study but they will not be dealt with in this paper, since our interest lies with differences in input. In what follows, we discuss studies which look at parents' behaviour and input as possible explanations for variability in children's language.

1.2. 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 steps, as well as in later steps (Bornstein et al. 1998, Florin 1999, Hart and Risley 1995, Huttenlocher et al. 1991, 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 (Goodman et al. 2008, Huttenlocher et al. 1991). Thus, in 16-month-olds, order of acquisition is closely linked with the relative frequency of words acquired in parents' speech (Huttenlocher et al. 1991). Overall, the word and grammatical category heard more often should be learnt earlier by children (Choi and Gopnik 1995, Goodman et al. 2008, Tardif 1996). Together with quantity, quality is a possible source of variation which has been analysed in caregiver-child interaction, e.g. based on mothers' responsiveness and volubility (Vanormelingen and Gillis 2016) or on the pragmatic values of caregiver utterances (see e.g. 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, saliency, as well as frequency and pragmatic aspects had an impact, especially as regards 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, Goldfield 2000, Tardif et al. 1997), Italian (e.g. Caselli et al. 1995, Tardif et al. 1997), Spanish (Jackson-Maldonado et al. 1993), French (Bassano 2000, Parisse and Le Normand 2000, Poulin-Dubois et al. 1995), Hebrew (Maital et al. 2000) but these results could not be replicated with other languages like Korean and Mandarin, where conflicting evidence

has been found (see Gentner 1982, Tardif 1996, 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, emphasising 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 (Fernald and Morikawa 1993, Gopnik et al. 1996, Tamis-LeMonda et al. 1992).

More recently, variation has been shown to depend more on context than language typology: in Altınkamış 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 analyses, consists in starting from context to analyse how it relates to language acquisition.

1.3. 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. Bruner underlined the importance of what he called formats in social interaction and showed that they are a crucial step 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 evidencing SES- or community-related differences (e.g. as regards the part played by parents' free time and dedication to children's development; and notably time spent reading (Weizman and Snow 2001): and 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 (Choi 2000, Raikes et al. 2006, Reese and Cox 1999), or free play (Choi 2000, Newland et al. 2001). The variety of contexts (Hoff-Ginsberg, 1991) as well as duration of caregiver-child interaction (Hoff et al. 2002, Snow et al. 1982) have been shown to have an impact. This line of research also related certain kinds of more specific activities to more advanced language abilities like literacy.

Here 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 with 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 quality, stability, engagement and sensitivity of caretakers interacting with their children (interactional style) (Leyendecker et al. 1997b, Leyendecker et al. 1997a, Masur and Gleason 1980, Newland et al. 2001, Tamis-LeMonda et al. 1996, 2004, Yont et al. 2003). 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 (Hoff-Ginsberg 1991: 783, Lewis and Gregory 1987, Snow et al. 1976, Wiley et al. 1989). 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 (both from the point of view of lexical diversity and syntactic diversity), large-scale studies have evidenced a positive impact of the time spent reading (Weizman and Snow 2001) with no clear SES-related differences (Hindman et al. 2013). A final 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" (ibid.) 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: 793).

The above 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 -e.g. the nouns bias (Altınkamış et al. 2014, Choi 2000, Kern et al. 2012), or on mothers' conversational style type (Bornstein et al. 1992, Flynn and Masur 2007, Golinkoff et al. 2015, Haden and Fivush 1996, Kelly et al. 2015, Kloth et al. 1998) 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 (Kern et al. 2012).

A final line of research has taken activities into account with a view to contextualising vocabulary development within the Human Speech home corpus (Roy 2009, Roy et al. 2006b, 2006a, 2009, 2012, 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-linguistics behaviour into account. Brandon Roy (Roy 2014) has explored activities in terms of spatial and temporal distribution as well as words used. 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" (Roy et al. 2015 : 1). 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, pilot 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 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.

1.4. Our study

As shown in our analysis of previous literature, the notion of context where acquisition takes place has not often been analysed as such. And yet, we know that context has an impact: according to the variety of social activities, 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 steps 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. Besides, rather than focusing on linguistic features like morphology and saliency 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 - description

Our data comprise 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 by Kelly et al. (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 less studies on children acquiring dialectical Arabic (Zalami 2007). When such studies exist, they deal with either children's learning of formal and so-called "standard Arabic"–which is a later acquisition, or 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, 2006). Our three datasets are the Providence corpus (Demuth et al. 2006) for American English; the PREMS corpus (Kern et al. 2009) for Tunisian Arabic and the French OHLL corpus (Kern 2005). We used data from 2 children per language: one boy and one girl. All families are high SES. Birth rank was not controlled but the information was collected for each child and gathered in table 1 below in order to confront it to our observations. Participants are not present 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 3 key linguistic steps only, i.e. 10, 50 and 100 words produced. The 10-word step was the most

homogeneous early step in our data and it corresponds to a well-described milestone in language development (Tardif et al. 2008), and so did the 50- and 100-word steps which precede and follow the lexical spurt (Benedict 1979, Bloom 1973, Fenson et al. 1993, Huttenlocher et al. 1991, Kern 2010). Thus we worked on 2.5 hours of recording for each child and 5 hours for each linguistic step, amounting to a total of 15 hours of recording.

Table 1: Corpus description

Children	Gender	Birth order	Language	Age in month for linguistic step			Time session / linguistic step			Total (hours)
				10 words	50 words	100 words	10 words	50 words	100 words	
US girl	F	2	American English	14;00	18;18	20;22	50 min	50 min	50 min	2.5
US boy	M	4	American English	16;28	19;28	21;08	50 min	50 min	50 min	2.5
French girl	F	1	French	15;26	19;29	21;20	50 min	50 min	50 min	2.5
French boy	M	2	French	14;29	18;14	19;4	50 min	50 min	50 min	2.5
Tunisian girl	F	2	Tunisian Arabic	17;1	21;3	22;21	50 min	50 min	50 min	2.5
Tunisian boy	M	1	Tunisian Arabic	14;16	21;22	25;6	50 min	50 min	50 min	2.5
Means & Totals				15.56	19.97	21.78	5 hours	5 hours	5 hours	15

In order to obtain equivalent linguistic steps while maximising comparability across datasets, we calculated cumulated types across all recordings, so that each linguistic step 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.

Table 2: Children's linguistic measures at each step

Child	Gender	Age	Overall number of utterances	Number of meaningful utterances	Number of utterances containing words	Cumulated type scores	Child number of types in recording	Child number of token in recording	Child MLU
10-word step									
American boy	M	16;28	141	20	18	7	7	18	
American girl	F	14;00	88	8	6	7	7	10	
French boy	M	14;29	418	17	14	6	6	15	
French girl	F	15;26	604	22	16	4	4	16	
Tunisian boy	M	14;16	na	13	13	10	10	15	
Tunisian girl	F	17;1	na	47	45	30	30	49	
Mean			312.75	21.17	18.67	10.67	10.67	20.5	
standard deviation			242.15	13.61	13.53	9.64	9.67	14.21	
50-word step									
American boy	M	16;28	282	95	95	40	35	157	1.9
American girl	F	14;00	317	76	76	32	27	121	1.87
French boy	M	14;29	487	112	92	33	28	99	1.36
French girl	F	15;26	375	156	105	30	27	145	1.49
Tunisian boy	M	14;16	na	47	28	19	19	31	na
Tunisian girl	F	17;1	na	142	136	86	86	172	1.27
mean			365.25	103.2	88.67	40	37	120.83	1.57
standard deviation			89.77	45.44	35.78	23.54	24.54	51.1	0.29
100-word step									
American boy	M	16;28	418	131	131	63	37	229	2
American girl	F	14;00	357	162	158	123	99	247	1.96
French boy	M	14;29	847	199	151	73	54	165	1.31
French girl	F	15;26	307	218	186	73	59	335	2.28
Tunisian boy	M	14;16	na	90	82	56	56	117	1.43
Tunisian girl	F	17;1	na	206	179	99	99	232	1.3
mean			482.25	167.67	147.83	81.17	67.33	220.83	1.71
standard deviation			247.37	49.79	37.84	25.16	25.7	74.58	0.42

As expected, when looking at those measures no child stands out across all steps. Nevertheless, for the 100-word step girls systematically over score boys in word type count in each of our three linguistic groups –i.e. 99 vs 37 for US children; 59 vs 32 for French children and 99 vs 50 for Tunisian children.

We computed mean length of utterance (MLU) from the second step onward, and more variation can be seen in mean length of utterance 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 mean length of utterances is linked with repeated utterances based on the same chunks[1]. As a result, the number of types remains relatively low. Besides, Tunisian mean length of utterances does not evolve 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 English or French. This did not have any impact on the comparability of our data, however, since types were retrieved and counted manually in order to establish cut-off points for our three developmental steps.

2.2. Transcription and coding

This section is mainly concerned with coding but since transcription implied decisions that were likely to have an impact on our analyses (especially as regards the linguistic measures), we start 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 to overall discourse quantity heard by a child (Weisleder and Fernald 2013). However, 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 wasn't supposed to take part in interactions.

2.2.1. Transcription

All main speakers in our recordings – i.e. at least child and mother -- were orthographically transcribed with PHON software using CHILDES rules (MacWhinney 2009).

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

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 words any linguistic forms 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 e.g. 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 nouns or verb, leaving out any other element in a third generic category.

In order to obtain comparable data, we adopted the coding categories used in previous studies (Choi 2000: 75). Thus, our noun category included proper names, kinship terms like *maman* and common nouns and excluded pronouns. Our category of verbs included main verbs and 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 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] However, these were not counted when calculating cumulated types to determine our key steps.

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. 1997b, Leyendecker et al. 1997b), activities were divided into 5 exclusive and exhaustive contexts: feeding, caretaking, toy play, social interaction and no interaction. Overall, functional and social contexts clearly moderated interactional experiences. SES effects on verbal and other interactional measures were limited to some 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 subjects from differing socioeconomic backgrounds and point to the need for further explorations of contexts.

In line with Bruner's notion of format and B.C. Roy's operationalisation of the notion as a core element in defining activity contexts, we coded the contexts of mother-child interactions. As opposed 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 drew a distinction between three main activities which occur regularly in any child's day: exploration, maintenance, and social activities. As shown in table 3 below, 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). We consider as maintenance activities, all activities revolving around food (dinners or snacks), health and hygiene (bath time, blowing the child's nose). Finally we consider as social activities, all activities of social play, book reading, and social speech (e.g. discussing previous experience).

Table 3: Coding categories of activities

Activity type	Activity subtype	Example
Solitary activities	Environment exploration	Play with washing machine

	Solitary play	Manipulation game alone
Maintenance	Feeding	Dinner, snaking, drink
	Cleanliness and hygiene	Diaper change, bath, hand whashing
Social activities	Social play	Play tea time, memories play, play hide and seek
	Book reading	Story or picture book reading by caregiver
	Social speech	Conversation with caregiver about the day, friend, pets ...

In what follows, we started by looking for regularities and variation across our three main types of activities but went back to our smaller scale (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 given language, and there were regular meetings to discuss coding and check reliability across languages.

Time spent interacting or doing things without speaking is not taken into account and not coded, so we don't 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 (Cartmill et al. 2013, Clark 2009). 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 step 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 step, for whole sessions, and depending on 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 VOCD 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 depending on activities. Proportions of nouns and verbs (tokens) per utterance were computed for each activity type, as was done in Choi's landmark study (Choi 2000: 78). Using per utterance ratios also helped us tone down the differences which resulted from our definition of word boundaries in Tunisian Arabic.

3. Analyses

Our exploratory 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 the different genders or birth-ranks come out. 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 6 children only, does not allow for any inferential statistics. Indeed, the probability that our tests result in type I error is stronger than that of finding any effect. In order to avoid overgeneralising 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 steps.

3.1.1. Global duration

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

Figure 1: all corpus activities

3.1.2. 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 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.

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

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 steps on activity types, apart from the French dataset where an increase of social activities and a corresponding decrease of solitary activities is 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, i.e. a third of the overall recording. The maintenance and solitary activities are more variable.

On the whole, rather than evidencing differences at each developmental steps, 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 are 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 steps. The first element to be taken into account as regards child-directed speech is the existence of disparities in the amount of speech addressed to the children in our recordings. We used utterance and word counts to assess these differences.

3.2.1.1. Utterances

As shown in Figure 3 below, 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.

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

The overall proportion of utterances in each activity context, language and step 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 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 step (especially in social context), so that mothers appear to be adjusting to their child's linguistic development.

3.2.1.2. Word tokens

While observed proportions of word tokens in CDS (Figure 4) are not considerably different from the above proportions of utterances, they evidence one trend which was less clear when looking at utterances only: the overall amount of speech addressed to girls is superior to

that which is directed to boys. It seems to be the case across languages, and more clearly so in the 100-word step.

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

3.2.1. 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): 50 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 step, and we had to pool all three linguistic steps together, as shown in Figure 5 below.

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 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.

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

3.2.2. 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 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 has 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's the Tunisian boy who receives the shortest utterances.

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

3.3. 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 Tariff'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 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 step 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.

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

4. Discussion

On the whole, this pilot study of activity contexts has evidenced differences in activity duration and distribution across step and language, and differences in child-directed speech that are more important across activity contexts than steps and languages. We did not, however, seek to achieve 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. **Erreur ! Source du renvoi introuvable.** below gives an overview of activity patterns by providing mean durations by language, gender (and therefore children and birth order) and activities.

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.

Table 4: Mean activity duration by language, gender, and activity

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

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 time of recording. Attention to this complexity is needed 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 implication 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 only be recorded 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. On top of 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 others 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 that are primarily focused on dyads such as ours we do not have only dyadic interactions. There are moments when the child is alone, other times with both parents or even with sister/ brother. In order to maximize comparability across datasets, we restricted our coding to dyads (child – parents), as is the case with most studies of child-directed speech. However, it is worth noting that as a result of this choice we left out 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 in 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 Violet. 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 and 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 moment of the day 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 -- e.g. 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 or caretaking, have been shown to have an influence on quantity of dyadic interaction observed (O'Brien and Nagle 1987, O'Brien et al. 1989) and can tone down certain differences which may be found across social classes as well as regarding the extent and nature of maternal discourse (Dunn et al. 1977, Hoff-Ginsberg 1991, Snow 1972, Wootton 1974). These elements are of interest to our results, since we found a lot of 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 enhanced by differences in SES[3]. Consequently, our study reinforces the claim that activity context should be taken into

[3] Our 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.

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 e.g. Gentner and Borodisky 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 and both American children, as is clear from Table 5 below. 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 –e.g. 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 only have 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.

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

Language	Gender	Activity	Mean
English	F	Social	5.61
English	F	Solitary	1.13
English	F	Maintenance	1.37
English	M	Social	9.40
English	M	Solitary	1.17
English	M	Maintenance	2.48

French	F	Social	5.72
French	F	Solitary	4.57
French	F	Maintenance	3.48
French	M	Social	5.65
French	M	Solitary	1.92
French	M	Maintenance	0.21
Tunisian	F	Social	5.80
Tunisian	F	Solitary	0.90
Tunisian	F	Maintenance	0.27
Tunisian	M	Social	3.12
Tunisian	M	Solitary	0.32
Tunisian	M	Maintenance	0.78

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

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

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

In order to account for this, we looked at the mothers' productions in maintenance contexts and found that in many situations (including e.g. 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.

Table 7: examples of utterances during maintenance activity

Maintenance (snacking) with English girl, 100-word step	
US girl's 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 step	
US boy's mother:	Wait a minute; we have some soup okay, next bite.
	Thank you, that's delicious.
	Thank you for sharing, have some soup now.
US boy:	No!
US boy's 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.

Another interesting feature of MLU is that it seems to vary according to the child's gender. On the whole, utterances directed to boys are shorter across all contexts, with smaller differences in maintenance contexts, and bigger differences in solitary contexts. This may be linked with differences in mothers' verbal 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 : 532).

Table 8: Mean Length Utterance by gender and activity

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

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 American caretakers than in any other language, and it was the case for the most part of our data. Overall, if previous studies have shown that in several linguistic communities mothers used more nouns than verbs (see e.g. Choi 1998, Fernald and Morikawa 1993, Goldfield 1993, Gopnik et al. 1996, Kim et al. 2000, Poulin-Dubois et al. 1995, Tardif et al. 1997) the reverse seems to be true for Tunisian Arabic, which to our knowledge was never 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 (Altınkamış et al. 2014; Choi 2000) 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, we will be able to look at subtypes and we may get more illuminating results for nouns and verbs.

4.1. 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 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 isolate 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 10 below gives examples of rare words found in our French data.

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

Craspouillou (funny nickname meaning dirty little one)
Mother comments upon child's activity while child explores environment alone **(solitary context)**
[qu'est+ce tu fais craspouillou là à baver ?]
(what are you doing here drooling around, dirty little one?)

Se déchaîner (to be frantic)
Child is overexcited and his mother comments upon his activity **(solitary context)**
[et mais tu te déchaîne toi ce matin, là hein.]
(you're frantic this morning huh)

S'extasier (to marvel)
The tower's rising and he will soon be able to destroy it **(social play with toy)**
[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)
Child began to be very exited **(social routine)**
[je te laisse là faire ton .] [ton sketch .]
(I'll let you do your …your comedy sketch)

Souk (mess or chaos)
Child moves all his toys out, mother comments upon his activity **(exploration context,)**
[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)
Mother comments on her child's **(solitary play)**
[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)

We sought to determine in which context our rare words appeared predominantly, so we computed means according to language and activity (table 11). Different means appear for

[44] with the exception of Tunisian Arabic where only three fully transcribed sessions were available.

Tunisian due to lack of transcribed data[5] 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 ones.

Table 10: Mean rare word percentage by language and activity

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

Looking at the detailed occurrences at each step (see Table 12 below), we see more occurrences of rare words in maintenance contexts at the 10-word step for English children and in solitary contexts again at the 10-word step for French children.

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

Language	Step	Activity	Mean
English	100 words	Social	0.08
English	100 words	Solitary	0.03
English	100 words	Maintenance	0.04
English	50 words	Social	0.06
English	50 words	Solitary	0.02
English	50 words	Maintenance	0.02
English	10 words	Social	0.06
English	10 words	Solitary	0.05
English	10 words	Maintenance	0.05
French	100 words	Social	0.04
French	100 words	Solitary	0.04
French	100 words	Maintenance	0.02
French	50 words	Social	0.01
French	50 words	Solitary	0.03
French	50 words	Maintenance	0.03
French	10 words	Social	0.02

[5] Thus, we could not outline rare words properly 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.

French	10 words	Solitary	0.05
French	10 words	Maintenance	0.01
Tunisian	100 words	Social	0.32
Tunisian	100 words	Solitary	0.21
Tunisian	100 words	Maintenance	0.30
Tunisian	50 words	Social	0.28
Tunisian	50 words	Solitary	0.20
Tunisian	50 words	Maintenance	0.27
Tunisian	10 words	Social	0.24
Tunisian	10 words	Solitary	0.19
Tunisian	10 words	Maintenance	0.14

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. And when we add all developmental steps 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 often appear 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.

4.2. Conclusion

The main finding of our pilot study is that apparent subcultural differences are maximized by focusing on single contexts and minimized by averaging across a variety of naturally-occurring contexts. Our study confirms that observation across a variety of uncontrolled activity contexts raises comparability issues (Lewedag et al. 1994, Lewis and Gregory 1987, O'Brien and Nagle 1987, Stevenson et al. 1986). 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 standardized situation (see e.g. Bornstein and Haynes 1998) and maybe 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.

References

Altınkamış, N. Feyza, Sophie Kern, und Hatice Sofu. 2014. When context matters more than language: Verb or noun in French and Turkish caregiver speech. *First Language* 34: 537–550.

Amayreh, Mousa M. 2003. Completion of the consonant inventory of Arabic. *Journal Of Speech, Language, And Hearing Research: JSLHR* 46: 517–529.

Au, T. K. F., M. Dapretto, und Y. K. Song. 1994. Input Vs Constraints: Early Word Acquisition in Korean and English. *Journal of Memory and Language* 33: 567–582.

Bassano, Dominique. 2000. Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language* 27: 521–559.

Bassano, Dominique. 1998. L'élaboration du lexique précoce chez l'enfant français : structure et variabilité. *Enfance* 51: 123–153.

Bates, Elizabeth et al. 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* 21: 85–123.

Bates, Elizabeth, Inge Bretherton, und Lynn Sebestyen Snyder. 1991. *From first words to grammar: Individual differences and dissociable mechanisms*. New York, NY, US: Cambridge University Press.

Bates, Elizabeth, und George F. Carnevale. 1993. New directions in research on language development. *Developmental Review* 13: 436–470.

Benedict, Helen. 1979. Early lexical development: Comprehension and production. *Journal of Child Language* 6: 183–200.

Bloom, Lois. 1973. *One word at a time: The use of single word utterances before syntax*. Mouton. The Hague, Pays-Bas.

Bornstein, Marc H. et al. 2004. Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American english. *Child Development* 75: 1115–1139.

Bornstein, Marc H. et al. 1992. Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development* 63: 808–821.

Bornstein, Marc H., Maurice O. Haynes, und Kathleen M. Painter. 1998. Sources of child vocabulary competence: a multivariate model. *Journal of Child Language* 25: 367–393.

Bornstein, Marc H., und O. Maurice Haynes. 1998. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development* 69: 654–671.

Brown, Roger. 1973. Development of the first language in the human species. *American Psychologist* 28: 97–106.

Bruner, Jerome. 1981. The social context of language acquisition. *Language & Communication* 1: 155–178.

Cartmill, Erica A. et al. 2013. Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences* 110: 11278–11283.

Caselli, Cristina, Paola Casadio, und Elizabeth Bates. 1999. A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language* 26: 69–111.

Caselli, Maria Cristina et al. 1995. A cross-linguistic study of early lexical development. *Cognitive Development* 10: 159–199.

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: 69–96.

Choi, Soonja. 1998. Verbs in early lexical and syntactic development in Korean. *Linguistics* 36: 755–780.

Choi, Soonja, und Alison Gopnik. 1995. Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language* 22: 497–529.

Clark, Eve V. 2009. *First language acquisition*. Cambridge, UK: Cambridge University Press.

Demuth, Katherine, Jennifer Culbertson, und Jennifer Alter. 2006. Word-minimality, Epenthesis and Coda Licensing in the Early Acquisition of English. *Language and Speech* 49: 137–173.

Dunn, Judy, Carol Wooding, und Joan Hermann. 1977. Mothers' speech to young children: Variation in context. *Developmental Medicine & Child Neurology* 19: 629–638.

Ely, Richard, Jean Berko Gleason, Ann MacGibbon, und Elena Zaretsky. 2001. Attention to Language: Lessons Learned at the Dinner Table. *Social Development* 10: 355–373.

Eriksson, Mårten et al. 2012. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology* 30: 326–343.

Farran, Dale C., und Ron Haskins. 1980. Reciprocal influence in the social interactions of mothers and three-year-old children from different socioeconomic backgrounds. *Child Development* 51: 780–791.

Fenson, Larry et al. 1993. *MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, US: Singular Publishing Group San Diego, CA.

Fenson, Larry et al. 1994. Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development* 59: 185.

Fernald, Anne, und Hiromi Morikawa. 1993. Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development* 64: 637–656.

Florin, Agnès. 1999. *Le développement du langage*. Paris, France: Dunod, DL 1999.

Flynn, Valerie, und Elise Frank Masur. 2007. Characteristics of maternal verbal style: Responsiveness and directiveness in two natural contexts. *Journal of Child Language* 34: 519–543.

Gentner, Dedre. 1982. *Why nouns are learned before verbs : linguistic relativity versus natural partitioning*. http://hdl.handle.net/2142/17514.

Gentner, Dedre, und Lera Borodisky. 2001. Individuation, relativity, and early word learning. In *Language Acquisition and Conceptual Development*, Hrsg. Melissa Bowerman und Stephen C. Levinson, 215–256. Cambridge, UK: Cambridge university press.

Goldfield, Beverly A. 1993. Noun bias in maternal speech to one-year-olds. *Journal of Child Language* 20: 85–99.

Goldfield, Beverly A. 2000. Nouns before verbs in comprehension vs. production: The view from pragmatics. *Journal of Child Language* 27: 501–520.

Golinkoff, Roberta Michnick, Dilara Deniz Can, Melanie Soderstrom, und 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: 339–344.

Goodman, Judith C., Philip S. Dale, und Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35: 515–531.

Gopnik, Alison, und Soonja Choi. 1990. Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language* 10: 199–215.

Gopnik, Alison, Soonja Choi, und Therese Baumberger. 1996. Cross-linguistic differences in early semantic and cognitive development. *Cognitive Development* 11: 197–225.

Haden, Catherine A., und Robyn Fivush. 1996. Contextual variation in maternal conservational styles. *Merrill-Palmer Quarterly* 42: 200–227.

Hamilton, Antonia, Kim Plunkett, und Graham Schafer. 2000. Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language* 27: 689–705.

Hart, Betty, und Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, US: Paul H Brookes Publishing.

Hayes, Donald P., und Margaret G. Ahrens. 1988. Vocabulary simplification for children: a special case of ‘motherese’? *Journal of Child Language* 15: 395–410.

Hindman, Annemarie H., Lori E. Skibbe, und Tricia D. Foster. 2013. 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: 287–313.

Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26: 55–88.

Hoff, Erika, Brett Laursen, Twila Tardif, und Marc H. Bornstein. 2002. Socioeconomic status and parenting. In *Handbook of parenting: Vol. 2: Biology and ecology of parenting (2nd ed.)*, 231–252. Mahwah, US: Lawrence Erlbaum Associates.

Hoff-Ginsberg, Erika. 1991. Mother-Child Conversation in Different Social Classes and Communicative Settings. *Child Development* 62: 782–796.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, und Thomas Lyons. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology* 27: 236–248.

Jackson-Maldonado, Donna, Donna Thal, Virginia Marchman, Elizabeth Bates, und Vera Gutierrez-Clellen. 1993. Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language* 20: 523–549.

Kauschke, Christina, und 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: 735–757.

Kelly, Barbara F., William Forshaw, Rachel Nordlinger, und Gillian Wigglesworth. 2015. Linguistic diversity in first language acquisition research: Moving beyond the challenges. *First Language* 35: 286–304.

Kern, S., B. L. Davis, und I Zinc. 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*, Hrsg. Francesco d’Errico und Jean-Marie Hombert, 205–332. Amsterdam/Philadelphia.

Kern, Sophie. 2005. De l’universalité et des spécificités du développement langagier précoce. In *Aux origines des langues et du langage*, Hrsg. JM Hombert, 270–291.

Kern, Sophie. 2010. Les premiers mots du jeune entant français: Analyse quantitative et qualitative du vocabulaire réceptif et productif des deux premières années de vie. *Rééducation orthophonique* 48: 149–165.

Kern, Sophie, Florence Chenu, und Feyza 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.

Kim, Mikyong, Karla K. McGREGOR, und Cynthia K. Thompson. 2000. Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. *Journal of Child Language* 27: 225–254.

Kloth, Saskia, Peggy Janssen, Floris Kraaimaat, und Gene J. Brutten. 1998. Communicative styles of mothers interacting with their preschool-age children: a factor analytic study. *Journal of Child Language* 25: 149–168.

Lewedag, Vanessa L., D. Kimbrough Oller, und Michael P. Lynch. 1994. Infants’ vocalization patterns across home and laboratory environments. *First Language* 14: 049-65.

Lewis, Charlie, und Susan Gregory. 1987. Parents’talk to their infants: the importance of context. *First Language* 7: 201–216.

Leyendecker, Birgit, Michael E. Lamb, und Axel Schölmerich. 1997a. Studying mother-infant interaction: The effects of context and length of observation in two subcultural groups. *Infant Behavior and Development* 20: 325–337.

Leyendecker, Birgit, Michael E. Lamb, Axel Scholmerichia, und 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: 15–34.

MacWhinney, Brian. 2009. *The CHILDES Project: Tools for Analyzing Talk*. 3. Aufl. Mahwah, US: Lawrence Erlbaum Associates.

Maital, Sharone L., Esther Dromi, Abraham Sagi, und Marc H. Bornstein. 2000. The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language* 27: 43–67.

Malvern, David, Brian Richards, Ngoni Chipere, und Pilar Durán. 2004. *Lexical diversity and language development*. New York, US: Palgrave Macmillan.

Masur, Elise F., und Jean B. Gleason. 1980. Parent–child interaction and the acquisition of lexical information during play. *Developmental Psychology* 16: 404–409.

Newland, Lisa A, Lori A Roggman, und Lisa K Boyce. 2001. The development of social toy play and language in infancy. *Infant Behavior and Development* 24: 1–25.

O'Brien, Marion, Jeanne M. Johnson, und Diana Anderson-Goetz. 1989. Evaluating quality in mother-infant interaction: Situational effects. *Infant Behavior and Development* 12: 451–464.

O'Brien, Marion, und Keith J. Nagle. 1987. Parents' speech to toddlers: the effect of play context. *Journal of Child Language* 14: 269–279.

Omar, Margaret K. 2007. *The Acquisition of Egyptian Arabic as a Native Language*. Georgetown University Press.

Pan, Barbara Alexander, Rivka Y. Perlmann, und Catherine E. Snow. 1999. Food for thought: Dinner table as a context for observing parent-child discourse. In *Methods for studying language production*, Hrsg. Lise Menn und Nan Bernstein Ratner, 205–222. Oxford, UK: Psychology Press.

Parisse, Christophe, und Marie-Thérèse Le Normand. 2000. How children build their morphosyntax: The case of French. *Journal of Child Language* 27: 267–292.

Poulin-Dubois, Diane, Susan Graham, und Lorrie Sippola. 1995. Early lexical development: the contribution of parental labelling and infants' categorization abilities. *Journal of Child Language* 22: 325–343.

Raikes, Helen et al. 2006. Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development* 77: 924–953.

Reese, Elaine, und Adell Cox. 1999. Quality of adult book reading affects children's emergent literacy. *Developmental Psychology* 35: 20–28.

Rondal, Jean A., Martine Ghiotto, Serge Bredart, und Jean-François Bachelet. 1987. Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language* 14: 433–446.

Rowe, Meredith L. 2008. Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language* 35: 185–205.

Roy, BC, MC Frank, und D Roy. 2012. Relating Activity Contexts to Early Word Learning in Dense Longitudinal Data.

Roy, Brandon C. 2014. „Précis of The Birth of a Word (summary)". Cambridge, US http://web.media.mit.edu/~bcroy/papers/bcroy-phd_precis.pdf.

Roy, Brandon C., Michael C. Frank, Philip DeCamp, Matthew Miller, und Deb Roy. 2015. Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences* 112: 12663–12668.

Roy, Brandon C., Michael C. Frank, und Deb Roy. 2009. Exploring word learning in a high-density longitudinal corpus. Amsterdam, The Netherlands: Cognitive Science Society.

Roy, Deb. 2009. New horizons in the study of child language acquisition. 13–20. Brighton, United Kingdom http://dspace.mit.edu/handle/1721.1/65900.

Roy, Deb et al. 2006a. The Human Speechome Project (book section). In *Symbol Grounding and Beyond*, Hrsg. Paul Vogt, Yuuya Sugita, Elio Tuci und Chrystopher Nehaniv, 192–196. Berlin, Germany: Springer Berlin Heidelberg.

Roy, Deb et al. 2006b. The Human Speechome Project (conf parper). Vancouver, Canada.

Silverman, Stacy, und Nan Bernstein Ratner. 2002. Measuring lexical diversity in children who stutter: Application of vocd. *Journal of Fluency Disorders* 27: 289–304.

Snow, Catherine E. et al. 1976. Mothers' speech in three social classes. *Journal of Psycholinguistic Research* 5: 1–20.

Snow, Catherine E. 1972. Mothers' Speech to Children Learning Language. *Child Development* 43: 549–565.

Snow, Catherine E., und Diane E. Beals. 2006. Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development* 2006: 51–66.

Snow, Chatherine E., Clara Dubber, und Akke De Blauw. 1982. Routines in mother-child interaction. In *The language of children reared in poverty: Implications for evaluation and intervention*, Hrsg. Lynne Vernon-Feagans und Dale Clark Farrar, 53–72. New York, Etats-Unis d'Amérique: Academic Press.

Stevenson, Marguerite B., Lewis A. Leavitt, Mary A. Roach, Robin S. Chapman, und Jon F. Miller. 1986. Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research* 15: 451–461.

Tamis-LeMonda, Catherine S., Marc H. Bornstein, Lisa Baumwell, und Amy Melstein Damast. 1996. Responsive Parenting in the Second Year: Specific Influences on Children's Language and Play. *Early Development and Parenting* 5: 173–183.

Tamis-LeMonda, Catherine S., Marc H. Bornstein, Lisa Cyphers, Sueko Toda, und Misako Ogino. 1992. Language and play at one year: A comparison of toddlers and mothers in the united states and japan. *International Journal of Behavioral Development* 15: 19–42.

Tamis-LeMonda, Catherine S., Jacqueline D. Shannon, Natasha J. Cabrera, und 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: 1806–1820.

Tardif, Twila et al. 2008. Baby's first 10 words. *Developmental Psychology* 44: 929–938.

Tardif, Twila. 1996. Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* 32: 492–504.

Tardif, Twila, Susan A. Gelman, und Fan Xu. 1999. Putting the "noun bias" in context: A comparison of english and mandarin. *Child Development* 70: 620–635.

Tardif, Twila, Marilyn Shatz, und 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: 535–565.

Thordardottir, Elin T., und Susan Ellis Weismer. 1996. Language assessment via parent report: development of a screening instrument for Icelandic children. *First Language* 16: 265–285.

Vanormelingen, Liesbeth, und Steven Gillis. 2016. The influence of socio-economic status on mothers' volubility and responsiveness in a monolingual Dutch-speaking sample. *First Language* 36: 140–156.

Vihman, Marilyn May, und Lorraine McCune. 1994. When is a word a word? *Journal of Child Language* 21: 517–542.

Weisleder, Adriana, und Anne Fernald. 2013. Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science* 24: 2143–2152.

Weizman, Zehava Oz, und Catherine E. Snow. 2001. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology* 37: 265–279.

Wiley, Angela R., C. shore, und W.E. Dixon. 1989. Situational Differences in the Type of Utterances Mothers Use with Thirteen-Month-Old Children. (poster). 21. Kansas Gity, MO. http://eric.ed.gov/?id=ED309854.

Wootton, A. J. 1974. Talk in the homes of young children. *Sociology* 8: 277–295.

Yont, Kristine M., Catherine E. Snow, und 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: 435–454.

Zalami, F.B. 2007. First language acquisition. In *Encyclopedia of Arabic Language and Linguistics*, vol. 2, 96–104. Boston.

Appendix 1: examples of activities in French corpus (French girl at 100-word step)

example of **exploratory activity** (solitary play)

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

example of **exploratory activity** (exploration)

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

example of **social activity** (manners routine)

21	CHI	[yyy .] – yyy.
22	MOT	[tu nous dis bonjour ?] – You say hello?
23	MOT	[tu fais un petit coucou ?] – You do a little hello ?
24	MOT	[non ?] – No ?

example of **social activity** (teach routine)

68	MOT	[combien il+y+en+avait de fleur PRENOM ?] – How there were flower NAME ?
69	CHI	[a plus .] – Is no more.
70	MOT	[il+y+en+a plus ?] – There is no more?
71	MOT	[oui non il+y+en+a plus maintenant .] – Yes non now there is no more.

example of **maintenance activity** (toilet-training)

472	MOT	[tu veux le pot ?] – you want pot ?
473	CHI	[non .] – no.
474	MOT	[maman elle va te chercher le pot ?] – Mom she will get you the pot?

example of **maintenance activity** (bath routine)

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

Figures

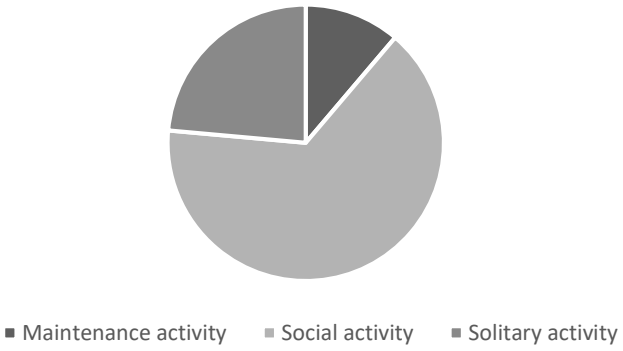


Figure 8: all corpus activities

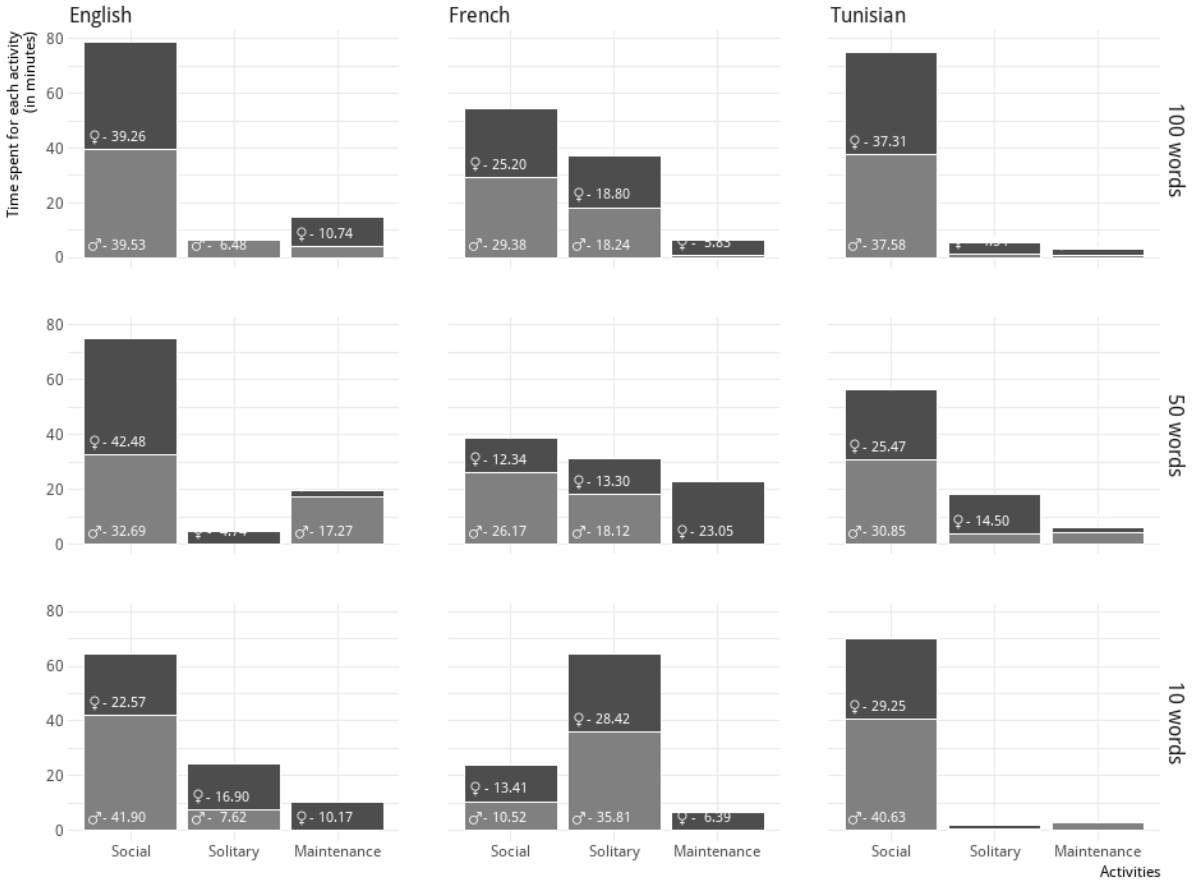


Figure 9: Activity duration per language, gender, and developmental step (labels corresponding to minutes spent)

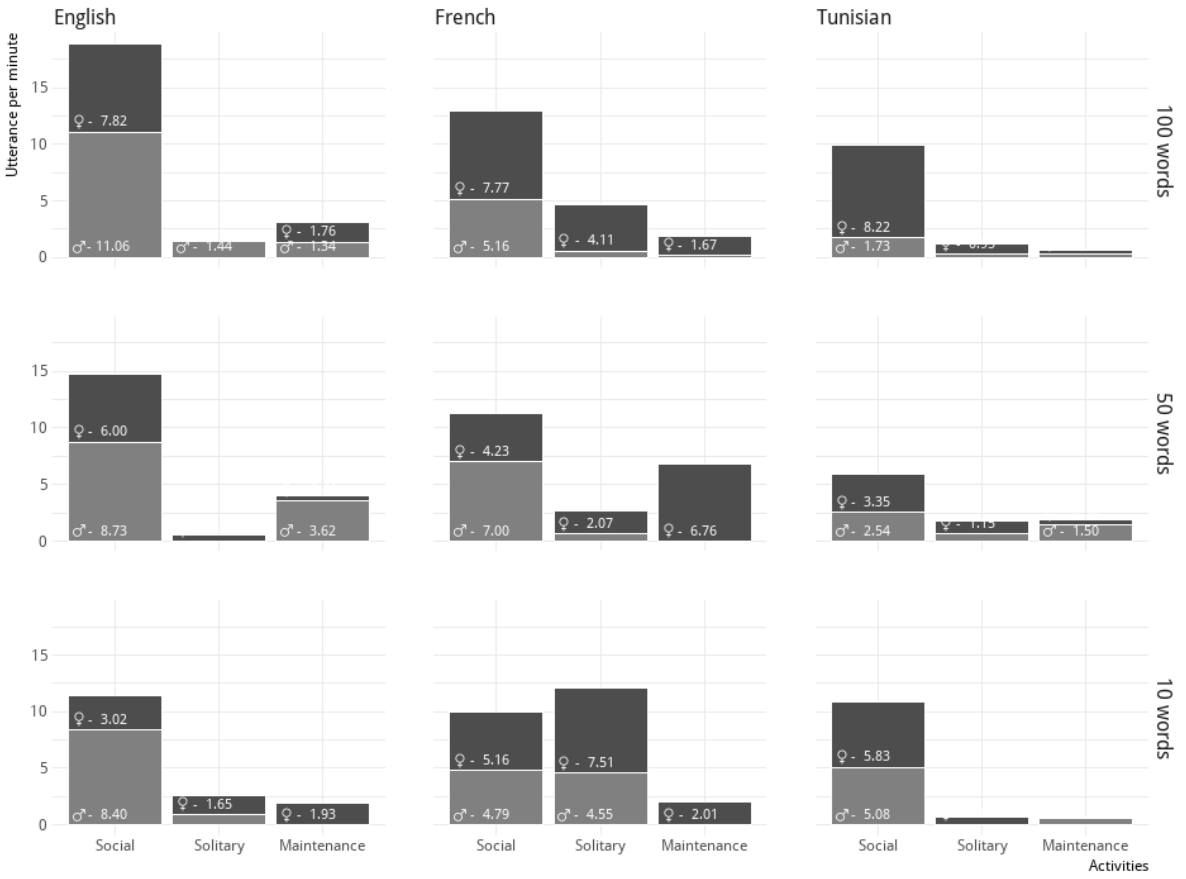


Figure 10: Utterances in child-directed speech per minute as a function of context and language

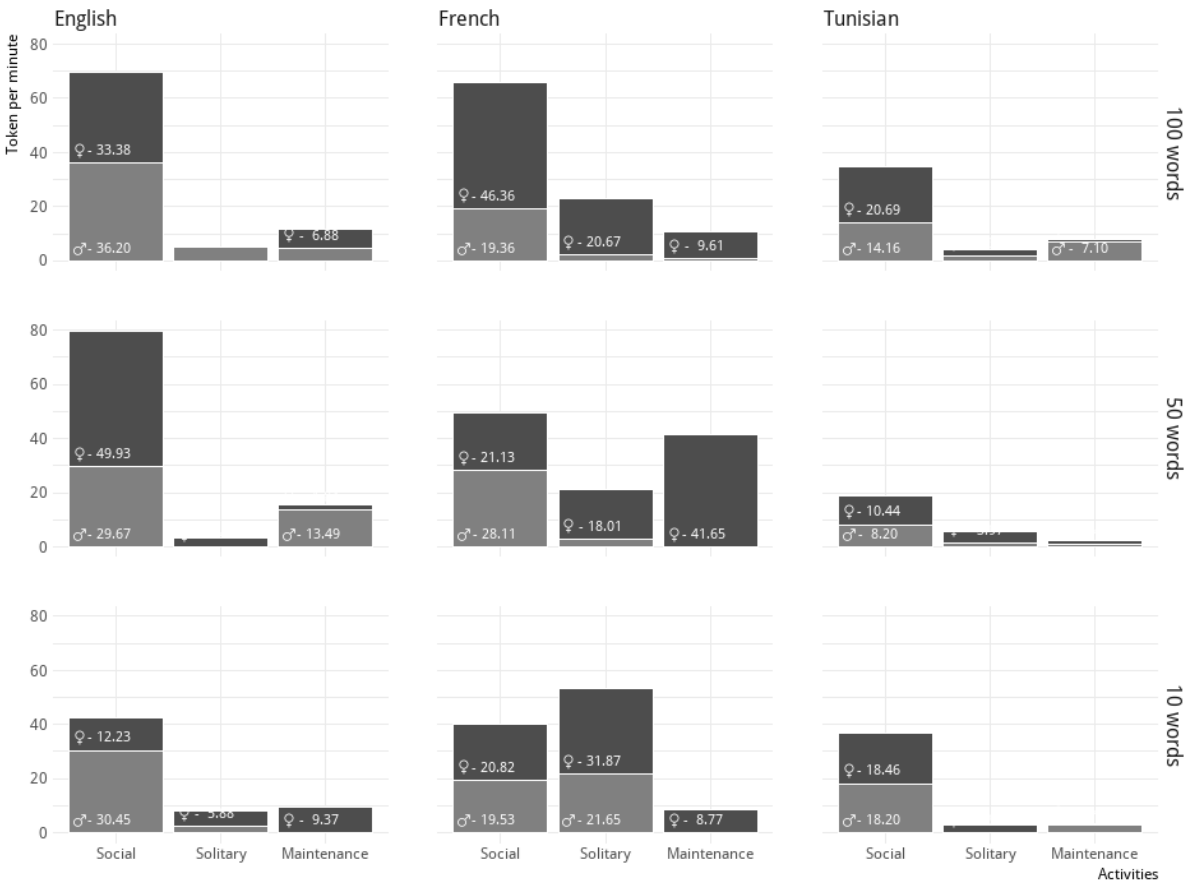


Figure 11: Word tokens in child-directed speech per minute as a function of context and language

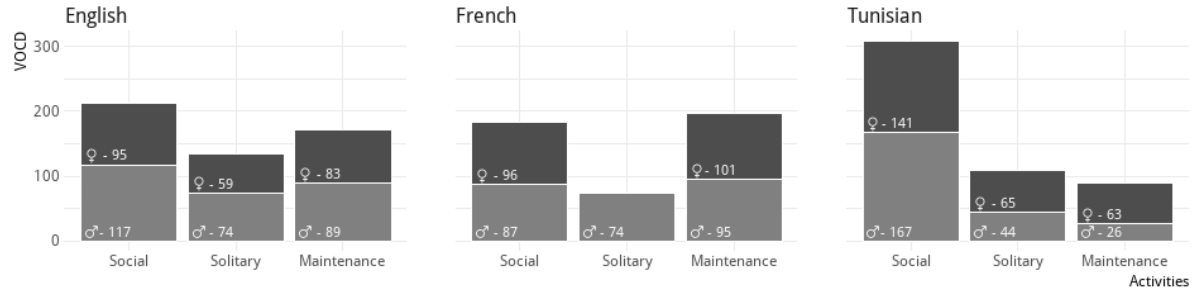


Figure 12: Lexical diversity (VOCD) in child-directed speech as a function of context and language (labels corresponding to different D values)

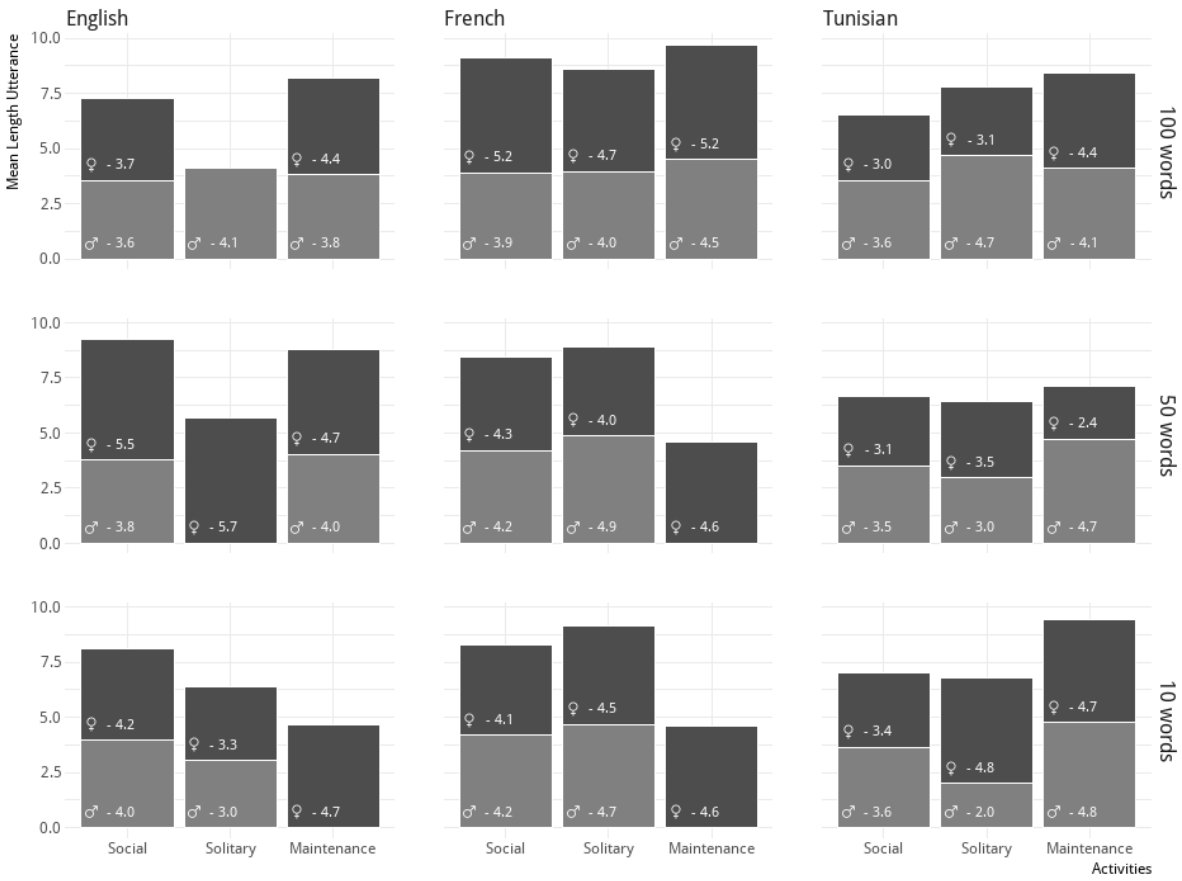


Figure 13: MLU in child-directed speech as a function of context and language (label corresponding to different MLU)

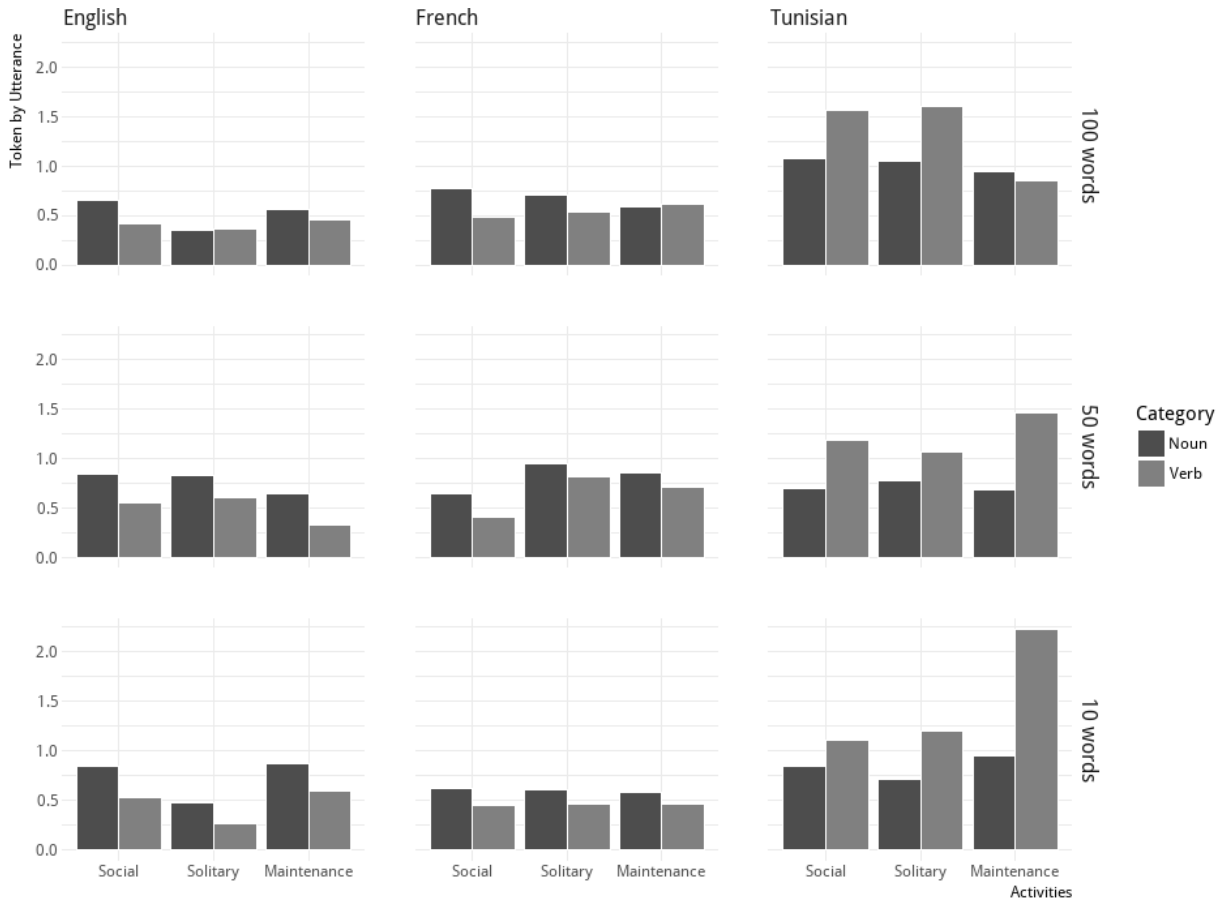


Figure 14: Nouns and verbs (tokens per utterance) in CDS as a function of context, language and developmental step