J. Child Lang. **35** (2008), 619–650. © 2008 Cambridge University Press
doi:10.1017/S0305000908008714 Printed in the United Kingdom

Early vocabulary development in Danish and other languages: A CDI-based comparison*

DORTHE BLESES

Center for Child Language, Institute of Language and Communication, University of Southern Denmark

WERNER VACH

Center for Child Language, Institute of Language and Communication, University of Southern Denmark and Department of Statistics, University of Southern Denmark

MALENE SLOTT

Center for Child Language, Institute of Language and Communication, University of Southern Denmark

SONJA WEHBERG

Department of Statistics, University of Southern Denmark and Institute of Language and Communication, University of Southern Denmark

PIA THOMSEN

Institute of Language and Communication, University of Southern Denmark

THOMAS O. MADSEN AND HANS BASBØLL

Center for Child Language, Institute of Language and Communication, University of Southern Denmark

(*Received* 11 *August* 2006. *Revised* 25 *April* 2007)

[*] We thank Larry Fenson and colleagues for permission to adapt the MacArthur-Bates Communicative Development Inventories to Danish. The data for the cross-sectional study presented in this paper were collected as part of a research project, 'Danish children's language acquisition' (2002–2005), funded by the Carlsberg Foundation and the University of Southern Denmark. Part of the data collection was funded by the Oticon Foundation. We also wish to thank the student assistants and all of the other people who have contributed to data collection and data entry. We also wish to thank all of the parents of the participating children who have completed CDI-parental reports for our studies. Finally, we want to express our thanks to Eva Sophia Myers, Rune N. Jørgensen, Kasper Østerholdt Jensen and Penny Kristiansen for their assistance in the preparation of this paper. Address for correspondence: Dorthe Bleses, Center for Child Language, Institute of Language and Communication, University of Southern Denmark, Denmark. Email: bleses@language.sdu.dk. Fax: +45 6550 3180.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

ABSTRACT

The main objective of this paper is to describe the trajectory of Danish children's early lexical development relative to other languages, by comparing a Danish study based on the Danish adaptation of *The MacArthur-Bates Communicative Development Inventories* (CDI) to 17 comparable CDI-studies. The second objective is to address the feasibility of cross-linguistic CDI-comparisons. The main finding is that the developmental trend of Danish children's early lexical development is similar to trends observed in other languages, yet the vocabulary comprehension score in the Danish children is the lowest across studies from age 1;0 onwards. We hypothesize that the delay is related to the nature of Danish sound structure, which presents Danish children with a harder task of segmentation. We conclude that CDI-studies are an important resource for cross-language studies, but reporting of studies needs to be standardized and the availability of published data improved in order to make comparisons more straightforward.

INTRODUCTION

Even though Denmark has fostered notable and internationally recognized linguists, including such famous personalities as Rasmus Rask (1787–1832), Otto Jespersen (1860–1943) and Louis Hjelmslev (1899–1965), research in child language acquisition in Denmark has been rather scattered up until recently and only relatively few studies on Danish children's early language acquisition have been carried out.

Otto Jespersen, in his 1916 book *Nutidssprog hos börn og voxne* [Modern language in children and adults], was in fact the first to publish empirical language data from Danish children based on a diary study. With the exception of a few studies in the 1970s it was not until the late 1980s/early 1990s that research in Danish children's language acquisition got off the ground, mostly as cross-linguistic ventures. Especially research on early LEXICAL development in Danish children has been very limited. To our knowledge, a total of only two studies has been published (cf. Brink, 1979; Plunkett, 1993), and until now no study has been carried out which has investigated the inter-individual variability in Danish children's early lexical development and which has established normative growth curves. Consequently, the question of how Danish children's early vocabulary development compares to children acquiring other languages has not been addressed.

Therefore, a large-scale cross-sectional study of Danish children's early language acquisition in the age range of 0;8 to 3;0 is the first to document the acquisition of early linguistic milestones in Danish children. The study is based on the Danish adaptation of the American CDI-instruments

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

(*The MacArthur-Bates Communicative Development Inventories*, cf. Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007). The Danish CDI-study is presented in Bleses *et al*. (in press; see also Bleses, Vach, Wehberg, Faber & Madsen, 2007). The main objective of this paper is therefore – based on data from the Danish study – to describe the trajectory of Danish children's early lexical development, as indexed by the CDI-parental reports, relative to other languages, by comparing the Danish study to all comparable CDI-studies in order to identify similarities and differences. The CDI-instrument (which consists of the two CDI-reports *CDI: Words and Gestures* and *CDI: Words and Sentences*; cf. Fenson *et al*., 2007) has not only been proven to be a very useful method for assessing (among other measures) young children's comprehension and production of early words for scientific as well as clinical purposes, but is also a very suitable instrument for cross-language comparisons. The American CDI-instrument has been adapted to numerous languages, permitting the unique possibility of controlled cross-linguistic comparisons between many different languages, as it is possible to compare parallel measures sampled using equivalent procedures across studies.

Findings from the American and the many other CDI-studies which have already been conducted (including typologically and genetically different languages, such as (several) Germanic, Romance, Slavic and Semitic languages), have so far shown remarkably similar patterns of acquisition and have identified far more similar than dissimilar tendencies characterizing the development of the lexicon for comprehension as well as production. According to Dale & Goodman (2005: 68), some of the main findings from English-speaking children which have been replicated in children speaking other languages are: (1) 'enormous variability in rate of lexical acquisition'; (2) 'a positive acceleration in the second year of life'; (3) 'comprehension and production are substantially dissociated with a ''fan effect'' relating them'; (4) 'gestural communication is more strongly related to receptive vocabulary than productive vocabulary'; and (5) 'vocabulary production is strongly related to grammatical development' (for specific comparisons across languages cf. e.g. Kovačević, Jelaska & Brozovic, 1998; Maital, Dromi, Sagi & Bornstein, 2000). Although most differences in developmental trends in various CDI-studies have been recognized in the grammar section across morphologically diverse languages (e.g. Devescovi, Caselli, Marchione, Pasqualetti, Reilly & Bates, 2005), some minor language-specific variations in lexical content and in the early composition of the lexicon have been identified, reflecting arbitrary relationships between sound and meaning across languages and cultural differences. Also, variation in the size of the lexicon, reflecting differences in morphology across languages, has been discussed (see e.g. Kovačević *et al*., 1998).

Adding early representative data from Danish children to this growing body of cross-linguistic CDI-data offers the opportunity of addressing

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

two hypotheses related to early lexical development, i.e. the role of sound structure and the role of 'early institutionalization'. First, Danish is very closely related, genetically as well as typologically, to the other Scandinavian languages. However, there is a crucial difference between Danish and the other Scandinavian languages, namely the particular reduction processes in Danish. In the words of Nina Grønnum, as a representative of many Danish scholars: 'An abundance of vowels, weak syllable codas, unstressed syllables without any vowel sound, and fairly inexpressive prosody makes Danish a harder nut to crack perceptually than most languages which it otherwise is reasonable to compare it to' (Grønnum 2003: 129). Studying vocabulary development in Danish children in comparison to the development in other languages may therefore increase our understanding of the role of sound structure for early vocabulary development (see below for a description of Danish sound structure). Second, the social situation of young Danish children differs in one respect from that of most other societies in the sense that the vast majority of Danish children attend daycare from an early age. Comparing this to the situation in other countries, the proportion of small children in formal childcare is rather high. Based on the OECD family database (www.oecd.org/els/social/family/database), covering the participation rates in formal daycare and preschool for children under the age of six in thirty OECD countries (based on figures from 2004), it appears that only Icelandic children under the age of three spend comparable time in daycare (61·6% of Danish children under the age of three are, according to the database, in formal daycare, compared to 58·7% of Icelandic children). In 70% of the countries, less than 30% of the children are in formal daycare, i.e. half of the children compared to the situation in Denmark (in eight countries, including Germany and Italy, 0–9·9% of the children are in formal daycare; in six countries the proportion is 10–19·9% of the children; in seven countries, including the US and UK, the proportion is 20–29·9% of the children; in four countries (including Sweden), the proportion is 30–39·9% of the children; and in one country the proportion is 40–50% of the children). After the first year of life, approximately 85% of the Danish children attend daycare, often between thirty-five and forty hours per week (cf. Bleses *et al.*, 2007). For these reasons, including Danish children in the CDI-comparisons will make it possible to study the potential implications of 'institutionalized' cultures such as Danish for language acquisition – a pattern which seems to become increasingly common in Western cultures.

In order to contrast the Danish CDI-study with other CDI-studies, we have compiled a list of all comparable studies based on available information. Using the CDI-website, published material and personal communications, it has been possible to identify forty-seven adaptations and eighteen studies (one of which is the Danish study) fulfilling the

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

following criteria: (1) they are based on comparable CDI-adaptations; (2) they are population based; and (3) they are (to some extent) published in manuals, papers or posters. To our knowledge, this paper is the first survey to compare early lexical development on the basis of all comparable CDI-studies (even if these comparisons are limited to selected analyses).

The second objective of the paper is methodological in nature. Here we want to address the question of comparability across CDI-studies with respect to the feasibility of cross-linguistic comparisons. How easy is it to collect information on CDI-studies? How comparable are CDI-adaptations? How many population-based studies can be found? How comparable are recruitment procedures? How are standard measurements reported? These are important questions, not least because one of the reasons for using adapted versions of the American CDI is the possibility of performing controlled language comparisons.

Initially, a short overview of the compilation of the CDI-studies, the structure of the adaptations and the sampling of children across studies are provided. The following analyses focus on identifying similarities and dissimilarities of the Danish children's lexical development relative to the other CDI-studies. This will be done mainly by comparing scores on the word lists of *CDI: Words and Gestures* and *CDI: Words and Sentences*, and in addition, supplementary analyses of other parts of *CDI: Words and Gestures* will also be included. We then discuss the findings, not least in the light of the particular perspectives actualized by the Danish study, and finally address the question of comparability across CDI-studies. However, before presenting the overview, we will provide a more detailed introduction to Danish sound structure.

Danish sound structure

Danish is rather unusual in having a complicated segmental phonology of reduction resulting from the combination of: (1) pervasive schwa reductions, in particular schwa assimilation; and (2) vocalizations of obstruents leading to alternations between obstruents and approximants or semi-vowels with the result that the inventories of segments in (a) onsets and (b) codas and interludes before neutral vowel phonemes, are radically different (see Bleses *et al.*, in press). The combination of these processes, in addition to weak prosodic cues, causes an indistinct syllable structure which in turn results in blurred vowel-consonant, syllable and word boundaries. In particular, word endings are often indistinctly pronounced, the final syllables of a word in many cases consisting only of monotonous vocalic stretches.

This indistinctness characterizing Danish sound structure is best illustrated by comparing Swedish and Danish. Take the words meaning

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

'street' and '(to) cook', respectively, in the two languages (broad transcription without diacritics, prosody not indicated).

		Swedish	Danish
(a)	'street'	*gata* [gɑːtɑ]	*gade* [gɛðˑ]
(b)	'to cook'	*koka* [kʰuːkɑ]	*koge* [kʰɔʊ]

The Swedish words for 'street' and 'cook' are *gata* and *koka*, respectively, which are pronounced with two distinct syllables. The two equivalent words in Danish are *gade* and *koge*, and they are instead pronounced without a clear two-syllable structure and without a clear syllable boundary. This is mainly due to the above-mentioned phenomena, i.e. the existence of the approximants or glides /ð/ and /ʊ/ instead of stops and the frequent reduction processes, as here the assimilation of schwa into adjacent sonorant segments (at http://www.sdu.dk/appendix you can find audio files with examples of Swedish and Danish words which illustrate this point). These processes do not only operate word internally, as can be seen in the following example (word prosodies, i.e. Swedish word tones and Danish stød, still not indicated).

		Swedish	Danish
(a)	'boil over'	*koka över* [kʰuːkɑœːvɐ]	*koge over* [kʰɔʊɒ̯ɐ]

In the Swedish word combination *koka över* the syllable boundaries are clearly marked. The equivalent word combination in Danish, *koge over*, is pronounced with an overlong rather monotonous vowel string after the initial consonant encompassing the whole next word, which consists of only a vowel string. An extreme instance is the (somewhat awkward) sentence *Det er hårdere at årelade* ('it is harder to bleed'), where the last three words are pronounced [hɒːːːlɛˑð], i.e. as one overlong vowel comprising five syllables, but (more moderate) examples of this vocalization process are pervasive in Danish. In brief, the vowel reduction processes therefore result in a vocalic monotonous sound structure, and in the discussion we will draw attention to the acquisitional implications of these processes.

METHOD

Procedure in the compilation of the CDI-studies

In order to enable controlled comparisons across the different languages, we wanted to include only those studies which are directly comparable to the Danish CDI-study. For this reason, a CDI-study must meet three criteria to be included: (1) the adaptation has to be as close as possible to the original American English CDI-reports; (2) the study has to be population based; and (3) some basic analyses of lexical development have to be

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

published (such as reported scores on basic parts of (at least) one of the two CDI-reports). By population based we mean that the study aims at collecting data from the general population and does not have a priori focus on a specific population, such as children with developmental delay, other particular health problems or bilingual children. We did not require that children were drawn randomly from a population register, but also accepted sampling techniques such as the recruitment of children from paediatricians, daycare centres or from other research programmes. In order to identify such studies, the following procedure was applied.

First, a list of CDI-adaptations was generated, starting with the list of adaptations available from the official CDI-homepage (http://www.sci.sdsu.edu/cdi/adaptations_ol.htm) and from personal communications. Forty-three languages (listed in the first column of Table 1) with a total of forty-seven CDI-adaptations were identified (the contact persons of all the adaptations are listed in the second column of Table 1). For English, French, German and Spanish, which are the official languages of different countries, different (regionally flavoured) adaptations have been developed (cf. Table 1 below), and these different adaptations are counted as different languages. Estonian, Greenlandic, Slovakian and Slovenian are not registered at the CDI-homepage.

As a second step, we started a search for information on all the CDI-adaptations. We classified all papers, books and other published material found as: (1) 'user's guides or technical manuals', i.e. reports (in most cases books) which include detailed information on the development of the Inventories (including validity and reliability measures) as well as detailed descriptive and correlational analyses of all the basic sections of the CDI-reports based on population-based norming studies (such as Fenson *et al.*, 2007); (2) 'basic papers', i.e. papers which also report descriptive and correlational analyses of (selected) basic sections of the CDI-reports (such as Hamilton, Plunkett & Schafer, 2000); and (3) 'other papers', i.e. papers which report other types of 'non-standard' analyses of the CDI-data (such as Fletcher, Chan, Wong, Stokes, Tardif & Leung (2004), which investigates the interface between phonetic and lexical abilities using the Cantonese CDI). Via email we also asked authors for permission to use results from posters which were presented at a 'European CDI poster session' at the IASCL (International Association of the Study of Child Language) conference in Berlin in 2005, mainly in order to include these data analyses in our comparisons in those cases where the results have not been published elsewhere (e.g. for Basque; and British English, the adaptation by Klee & Stokes, 2005). We found information for thirty-nine adaptations (cf. the third column of Table 1) counting thirty-five languages, and these adaptations have in all cases been used to conduct different kinds of studies.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

TABLE 1. *Overview of all CDI-adaptations and studies*

1	2	3	4	5	6	7
Language	Adaptation(s) (contact person)[6] (ip = in progress)	Info	Standard CDI	Standard analyses of population based studies published in manuals [M], basic papers [BP] or posters [PO]	Other studies/analyses published	Study incl.
Danish	Dorthe Bleses	x	x	M: Bleses *et al.* (2007); BP: Bleses *et al.* (this volume)	x	x
Bangladesh	Jena Hamadani	x	(short list)		Hamadani *et al.* (2006) [abstract]	
Basque	Margareta Almgren	x	x	PO: Almgren *et al.* (2005)		x
Catalan	Elisabet Serrat Sellabona (ip)	—				
Chinese, Mandarin	Twila Tardif	x	x	PhD thesis: Wu (1997)[1]	x	x
Chinese, Cantonese	Twila Tardif	x	x	[Manual 'in prep']7	Fletcher *et al.* (2004)	
Croatian	Melita Kovačević	x	x	BP: Kovačević *et al.* (1998)[2]		x
Dutch	Inge Zink	x	x	M: Zink & Lejaegere (2002)	x	x
English, American	Larry Fenson	x	x	M: Fenson *et al.* (2007)	x	x
English, British	1. Janette Atkinson	x	(shortened version)		Atkinson *et al.* (2001)	
	2. Antonia Hamilton	x	x [3] (W&G)	BP: Hamilton *et al.* (2000)		x
	3. Thomas Klee	x	x (W&S)	PO: Klee & Stokes (2005)		x
	4. Julian Pine	x	(x)[4]		Pine *et al.* (1996)	
English, New Zealand	Elaine Reese	x	x (W&S)		Reese & Read (2000)	
Estonian	Tiia Tulviste	x	x			
Finnish	Paula Lyytinen	x	x	M: Lyytinen (1999)	x	x
French, Canadian	Diane Poulin-Dubois	x	info not available		x	

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

French, European	Sophie Kern	x	x	BP: Hilaire *et al.* (2001); Kern (2007)		x
Galician	Miguel Pérez Pereira	x	x	BP: Pérez Pereira[2] *et al.* (2005)	x	x
German	1. Hannelore Grimm	x	(short list)		Grimm & Doil (2000)	
	2. Gisela Szagun	x	x	BP: Szagun & Steinbrink (2004)[5]		x
German, Austrian	Peter B. Marschik	x			Marschik *et al.* (2006)	
Greek	Ursula Stephany (ip)[7]	x				
Greenlandic	Ulla Carl (ip)[7]	x				
Hungarian	Leslie Nabors Oláh	—				
Hebrew	Sharone L. Maital	x	x (W&S)	BP: Maital *et al.* (2000)		x
Icelandic	Susan Ellis Weismer	x	x (W&S)	BP: Thordardottir & Weismer (1996)		x
Italian	Maria Cristina Caselli	x	x	M: Caselli & Casadio (1995)	x	x
Japanese	Tamiko Ogura	x	x		Ogura *et al.* (1993)	
Kiswahili & Kigiriama	Katie Alcock	x	info not available		x	
Korean	—	x	x		Kim *et al.* (2000)	
Malawian	Elizabeth Barcikowski (ip)	—				
Polish	Magdalena Smoczynska	x	x	[Norming study ip][7]		
Portuguese, Brazilian	Elizabeth Teixeira	x	x		Teixeira (2005)	
Romani	Hristo Kyuchukov (ip)	—				
Romanian	Elena Geangu	—				
Russian	Inna Chistovich	—				
Slovakian	Svetlana Kapalková (ip)	x				
Slovenian	Ljubica Marjanovic Umek	x	x (W&S)	[Norming study ip][7]		
Spanish, Cuban	Barbara Pearson	x	x		Fernández & Umbel (1991)	
Spanish, Mexican	Donna Jackson-Maldonado	x	x	M: Jackson-Maldonado *et al.* (2003)	x	x
Spanish, European	Susana López-Ornat	x	x	M: López-Ornat *et al.* (2005)	x	x
Swedish	Mårten Eriksson	x	x	M: Eriksson & Berglund (2002)	x	x
Tamil	Nitya Sethuraman (ip)	x				

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

Table 1. (*Cont.*)

1	2	3	4	5	6	7
Language	Adaptation(s) (contact person)[6] (ip = in progress)	Info	Standard CDI	Standard analyses of population based studies published in manuals [M], basic papers [BP] or posters [PO]	Other studies/analyses published	Study incl.
Thai	Sudaporn Luksaneeyanawin	—				
Turkish	Aylin Küntay (ip)	x				
Yiddish	Isabelle Barriere	—				

[1] On the official CDI-homepage, this study is placed under Chinese. Since the thesis concerns Mandarin Chinese it is listed here under Chinese, Mandarin.
[2] In our comparisons later, we also use results of two posters from the 'European CDI poster session' presented at IASCL, 2005, in Berlin by Cepanec *et al.* (Croatian) and Pérez Pereira *et al.* (Galician).
[3] Different forms of the CDI: *Words and Gestures* developed (only vocabulary list). The final Oxford version is developed in order to be as similar to the original American CDI: as possible.
[4] An additional multiword utterance checklist was added as well as two open sections.
[5] The adaptation is not quite compatible since the wordlist only contains 571 words (approximately 100 words were not adapted, as the authors found the words were too intertwined with the American culture).
[6] The contact persons are – with few exceptions – taken from the official CDI-homepage. Where more than one person was listed, only the first person is included in the table in order to save space.
[7] Personal communication.

On the basis of our review, twenty-eight adaptations were identified as fulfilling the criterion of a 'standard' adaptation (cf. the fourth column in Table 1). For two adaptations, the relevant information could not be found (i.e. Canadian French and Kiswahili & Kigiriama). Three adaptations only exist as a 'short form' (Bangladeshi, British English (the adaptation by Atkinson, Anker, Braddick, Nokes, Mason & Braddick, 2001) and German (the adaptation by Grimm & Doil, 2000)). The Austrian German adaptation has, in part, a different structure (Marschik *et al.*, 2006), and the Japanese adaptation (Ogura, Yamashita, Murase & Dale, 1993) combines *CDI: Words and Gestures* and *CDI: Words and Sentences*. Hence, all these adaptations could not be regarded as close enough to the original American CDI. Six adaptations (based on published material) cover only one of the two CDI-reports, but are included in the comparisons, together with some other adaptations showing minor differences, which we regarded as negligible for our purpose.

We identified twenty-one population based studies (cf. column five of Table 1). Three studies are not yet published (Chinese Cantonese, Polish and Slovenian), eight studies are published in manuals, seven studies are published in basic papers, one as a PhD thesis, and two studies have been presented only in posters (cf. Table 1). These eighteen published/presented studies (marked in the final column of Table 1), including the Danish study, will be the basis of the cross-linguistic comparison presented in this paper. For the sake of completeness we also include in column 6 of Table 1 the most relevant sources of our investigations for those adaptations where no population-based study could be identified. The scope of the studies included and the type of analyses which have been performed vary, as we shall see, quite extensively.

A comparison of the structure of the adaptations of the MacArthur-Bates Communicative Development Inventories

The structure of the adaptations of the two CDI-reports *CDI: Words and Gestures* and *CDI: Words and Sentences* used in the population-based studies are described here (information on the Danish adaptation can be found in Bleses *et al.*, in press). As one would expect, given that only studies which are based on a close adaptation to the original American English study are included, the structure of the adaptations – and the number of items – are quite similar. The number of word categories in *CDI: Words and Gestures* ranges from 19 to 22, whereas the number of words varies from 303 to 434, mainly due to assimilations to cultural and linguistic characteristics in each adaptation. In *CDI: Words and Sentences* the variation is somewhat larger, from 571 words in the German adaptation to 725 in the Danish adaptation, but the numbers of word categories

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

are comparable across adaptations. The grammar section of *CDI: Words and Sentences* is significantly different across the languages, reflecting the very different morphological systems according to language typologies represented in the studies included; from richly inflected languages such as Croatian and Hebrew to less inflected languages like English and Mandarin Chinese (as the extreme).

However, even though all the adaptations included are constructed so as to resemble the original American CDI as much as possible, linguistic and cultural modifications have been made in each adaptation. Consequently, there is a certain degree of variation between the adaptations, which needs to be taken into consideration when interpreting the results of any cross-linguistic comparison (not least for the grammar section of the CDI, cf. e.g. Kovačević *et al.*, 1998).

Sampling in the studies included in the cross-linguistic study

Table 2 specifies sampling procedure and population size by type of CDI-report for each of the CDI-studies.

As can be observed in Table 2, the studies differ substantially in terms of the sampling procedure, ranging from random selection from population registries (the Danish and Swedish studies), the employment of 'subject pools' (the American study) and other research programmes (the Hebrew study) to recruitment from paediatricians (the Italian and French studies) and daycare centres (the Basque and Croatian studies). Most studies used several procedures. Similarly, the population size differs substantially across studies from 30 children in the Mandarin Chinese study to 6,112 children in the Danish study.

Analyses and statistics of the cross-linguistic study

In preparing the analyses for the cross-linguistic comparisons, we first examined which kinds of scores were available across the studies included. Our main interest was in vocabulary scores (comprehension and production), but we also wanted – for methodological reasons and in order to be able to make tentative comparisons of lexical development with other areas of development – to include other types of scores which could inform us about potential differences across the studies. To this end, we completed a detailed overview of the different analyses of the basic parts in the two CDI-reports which were carried out in the studies included (the overview is available at http://www.sdu.dk/appendix).

Two observations need to be put forward. First of all, the number of measures reported varies substantially across studies, from the American English, Mexican Spanish, European Spanish and Italian studies,

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

TABLE 2. *Sampling procedure and population size by study*

		CDI: Words and Gestures (0;8–1;3)	*CDI: Words and Sentences* (1;4–2;6)
Language (adaptation)	Sampling procedure	# of children*	# of children*
Danish	Central population registry: mail, telephone	2398 (0;8–1;8)	3714 (1;4–3;0)
Swedish	Central population registry: mail, telephone	228[1]	336[1]
Dutch	No information available	1024 (0;8–1;4·5)	958
English, American	Birth announcements, research subject pool, mailing list. From local doctors' offices and daycare programmes	659 (0;8–1;4)	1130
English, British (Hamilton *et al.*)	Parental reports were collected when children came to participate in tests at Oxford BabyLab; children were recruited from birth departments at hospitals	669 (1;0–2;1)	—
English, British (Klee & Stokes)	Mail	—	303
Finnish		95[2] (1;0+1;2)[3]	95[2] (1;6+2;0+2;6)[3]
German (Szagun & Steinbrink)	Recruitment from paediatricians	—	333 (1;6–2;6)
Icelandic	Public and private daycare, church parent meetings	—	72 (1;3–3;2)
French	Questionnaires distributed by paediatricians during a visit	548 (0;8–1;4)	663
Galician	Recruitment from other programmes	231	492
Italian	Paediatricians and staff members within the national health-care system	315 (0;8–1;5)	386
Spanish, European	Paediatricians, daycare	412	593
Spanish, Mexican	Advertisements in newspapers, recommendations from paediatricians; data collectors, who contacted parents through relatives, churches, social programmes, etc.	778 (0;8–1;6)	1094
Croatian	Daycare	250 (0;8–1;4)	349
Hebrew	Recruitment from a longitudinal early childcare study	—	253 (1;6–2;0)
Basque	Daycare and direct contact with families	468 (0;8–1;4)	972
Chinese, Mandarin	Chinese School in Boulder	15 (0;8–1;6)	15 (1;7–2;6)

* Information on population size is based on the papers listed in Table 1. Information concerning the Croatian data is based on Cepanec *et al.* (2005).
[1] Data is available every second month from 0;8 to 2;10.
[2] It is 95 children in all.
[3] Data is only available for five different time points.

which – not surprisingly, as these were the first studies to emerge – report scores in many different formats of all parts of the CDI, to studies like the British English one based on the Hamilton *et al.* (2000) adaptation and the Dutch study, which report only vocabulary scores. Besides these in-depth studies (including the Swedish study), other publications tend to focus on reporting results on the composition of the early vocabulary (e.g. the Italian, Hebrew and Mexican Spanish studies) and the relationship between lexical and grammatical development. All of the studies included in our cross-linguistic comparison did – in some way – report at least on one score for vocabulary development, which is the most reported part of the CDI-instrument.

This leads on to the next observation. The way in which scores are reported is likewise very diverse, varying from a variety of different statistics (in particular for the languages for which a manual has been published, except for the Dutch and the Finnish manuals) to only means. Owing to these divergences, the possibility for making controlled comparisons across the population-based studies included in the cross-linguistic study is limited. In fact, it was not possible to uncover a single statistic used by all studies for the same measure, not even for vocabulary comprehension and production. Accordingly, no comparison can include all studies.

Table 3 lists the analyses that we – on the basis of the options listed in the overview at http://www.sdu.dk/cfb – decided to include in our cross-linguistic study. Apart from vocabulary scores (comprehension as well as production) in the age range 0;8 to 2;6, we included analyses of section 1B, 'Phrases' (which are directly related to the early comprehension of words), and section 1C, 'Starting to talk' (which are directly related to the early production of words). Table 3 also indicates, for each measurement, which studies are included in the comparisons.

For some of the studies, not all statistics used subsequently in our analyses were directly available. The Dutch study reported only gender-specific medians for both word comprehension and production. We averaged the two values to obtain (a close approximation to) the population median. The Swedish study includes only children with an age at even months, so we used interpolation to obtain values for odd months (cf. Eriksson & Berglund, 2002). The Finnish study reports only data from a limited number of months and here we also used interpolation. The Italian data are in all cases read from graphs (in which the data has been summarized in age groups of two months). In Figure 1, the British English and Croatian data are read from graphs. In Figure 4, the British English data are read from a graph. Moreover, the median values on the production score of the *CDI: Words and Gestures* in the Croatian study are based on an interpolation of results presented in Table 3 of Kovačević *et al.* (1998). In Figure 7, the

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

TABLE 3. *Overview of the types of analyses included in the cross-linguistic study by type of score and language*

	Danish	Swedish	Dutch	English, American	English, British (Hamilton *et al.*)	English, British (Klee & Stokes)	German	Icelandic	Finnish	French	Galician	Italian	Spanish, European	Spanish, Mexican	Croatian	Hebrew	Basque	Chinese, Mandarin
CDI: Words and Gestures																		
B: Phrases																		
Phrases (median)	x	x	x	x					x	x	x		x	x				
Phrases (mean)												x						x
C: First sign of understanding																		
Imitation (mean)	x	x		x						x			x	x				(x)
Labelling (mean)	x	x		x						x			x	x				(x)
D: Vocabulary checklist																		
Word comprehension (median)	x	x	x	x	x				x	x	x	x	x	x	x		x	
Word comprehension (percentiles)	x	x																
Word comprehension (mean)																		x
Word production (median)	x	x	x	x	x				x	x	x	x	x	x	x		x	
Word production (mean)																		x
CDI: Words and Sentences																		
A: Vocabulary checklist																		
Word production (median)	x	x	x	x		x			x	x		x	x	x	x	x		
Word production (mean)							x	x										x

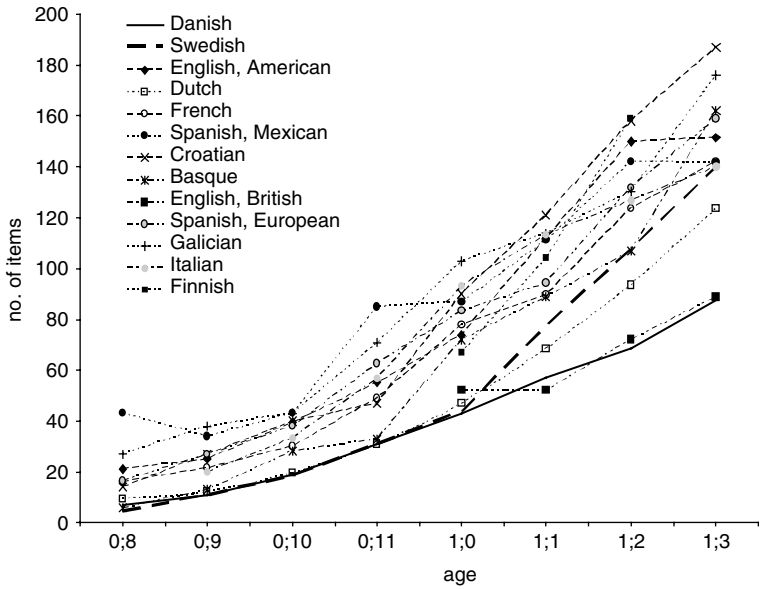


Fig. 1. Median vocabulary comprehension scores by age and study (*CDI*: *Words and Gestures*).
NOTES: The Mandarin Chinese study only reports mean values for one age group: 0;7–1;0: 29·40. The differences between the studies were significant ($p < 0{\cdot}0001$).

British English, European French and Hebrew data are read from graphs. The Mandarin Chinese, German and Icelandic studies report only mean values. The Italian study reports only means for comprehended phrases (cf. Figure 3). These values are given below the relevant figures.

In all figures the Danish study is marked with a solid black line. The significance of differences in age profiles among all studies is assessed by a Friedman test, excluding those studies which do not cover the whole age range. We do not report *p*-values for specific findings like differences in certain time ranges or changes in the degree of increase, as these findings have to be regarded as post hoc hypotheses developed at the hands of our data.

RESULTS

We will initially present global comparisons of the age-related changes in the two vocabulary scores of *CDI: Words and Gestures* (the comprehension score and the production score), examining similarities and differences across studies, with the main focus on identifying the developmental trends of Danish children's early lexical development relative to the other languages included in the comparisons.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

First, the median values for the comprehension score by age are plotted in Figure 1 for each study (language). The comparison is based on fourteen studies, including the Danish study. The variation of median score in the studies spans a difference of 5 words (Swedish) to 50 words (Mexican Spanish) at age 0;8 to a difference between 87 (Danish) to 187 words (Croatian) at age 1;3. Between 0;10 and 1;0, the curves in practically all of the studies suddenly become steeper, indicating some kind of vocabulary spurt in comprehension. However, no such acceleration in the growth rate occurs in the Danish data. Rather, a simple linear trend can be observed here over the whole period, implying that the Danish children produce the lowest score from age 1;0 onwards. A similar pattern can be observed only in the British English-speaking children, who have an almost equally low score (at age 1;1, the Danish children even produce a few words more than the British English children).

From a general point of view the developmental profiles across studies are remarkably similar and this confirms previous anticipations, even though the variation in the median score across studies within this age range is extensive, with the largest difference being 100 words (between the Danish and the Croatian data at age 1;3). The developmental profile of the comprehension vocabulary in the Danish children, however, is somewhat different than what can be observed in the other studies, and the score is relatively and remarkably low by comparison. Most surprising, perhaps, are the divergent acquisitional patterns of the Danish and the Swedish children, considering the typological closeness between these languages (Swedish is the most closely related to Danish among all the languages included). Here we can observe a distinct increase in the Swedish children starting at age 1;0, which is absent in the Danish children in Figure 1 (actually, as the Danish study includes children up to the age of 1;8, an increase can be observed at the age of 1;3 (see Bleses *et al.*, in press). It is therefore interesting to take a closer look at the Danish and Swedish data.

Figure 2 shows the variation in vocabulary size in comprehension for Danish and Swedish children stratified by age by presenting selected percentiles of the comprehension score. The median and percentile scores (10th, 25th, 75th and 90th percentile) are represented by different symbols in the two languages. Figure 2 reveals that Danish children are not delayed compared to Swedish children at the age of 0;8. There is not only a slight advantage in the median compared to the Swedish children, but a much more distinct advantage in the higher percentiles. However, already at age 0;11 we can no longer observe an advantage for any percentiles. A delay of the Danish children becomes visible earlier in the higher percentiles, and in general the difference in the lower percentiles is smaller than the difference in the higher percentiles for fixed ages. The largest difference across the two

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

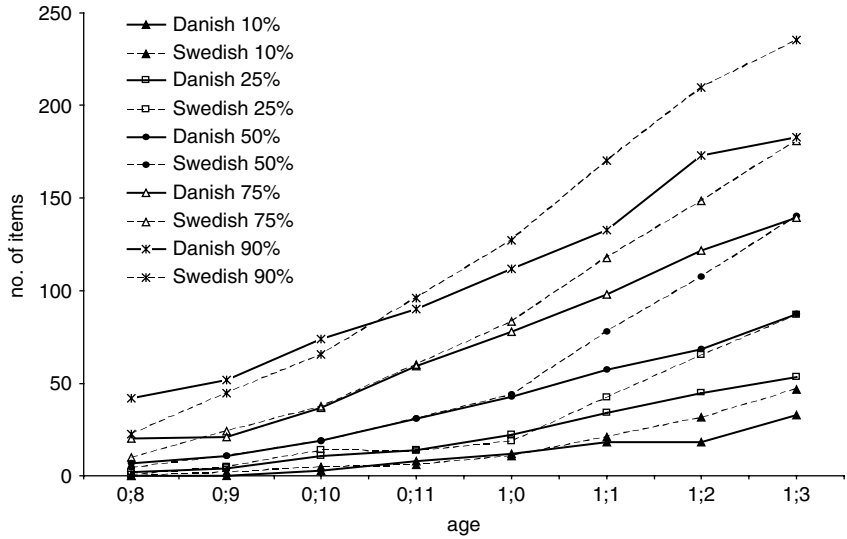


Fig. 2. Variation in comprehension scores in the Danish and Swedish studies (*CDI: Words and Gestures*).

data-sets is at age 1;3 where the median and 90th percentile score in both cases differ by 52 words. Another important finding is that the rather flat developmental curve of the Danish children in comparison with the Swedish children observed in Figure 2 is evident for all percentiles, suggesting that this is a rather robust finding. Summing up, in particular the better half of the Danish children (i.e. those above the median) are lagging behind from age 1;0.

In order to shed more light on the early comprehension abilities of the Danish children we now turn to another more indirect measure of word comprehension in *CDI: Words and Gestures*, the comprehension of phrases (1B), which is given in Figure 3. Eleven studies (including the Danish study) report data from this section of the CDI.

The median values of the number of comprehended phrases for the Danish children reveal a developmental trend in the age span covered by the CDI-report (age 0;8: 2 phrases to age 1;3: 16 phrases), but does not approach ceiling (26 phrases). The Danish children score lower at all times when compared to all of the other languages included in the comparison, and furthermore the gap seems to increase after age 1;0. This adds further support to the finding that the early comprehension in Danish children is to some extent slower, compared to the early comprehension in children speaking the languages represented in the population-based CDI-studies.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

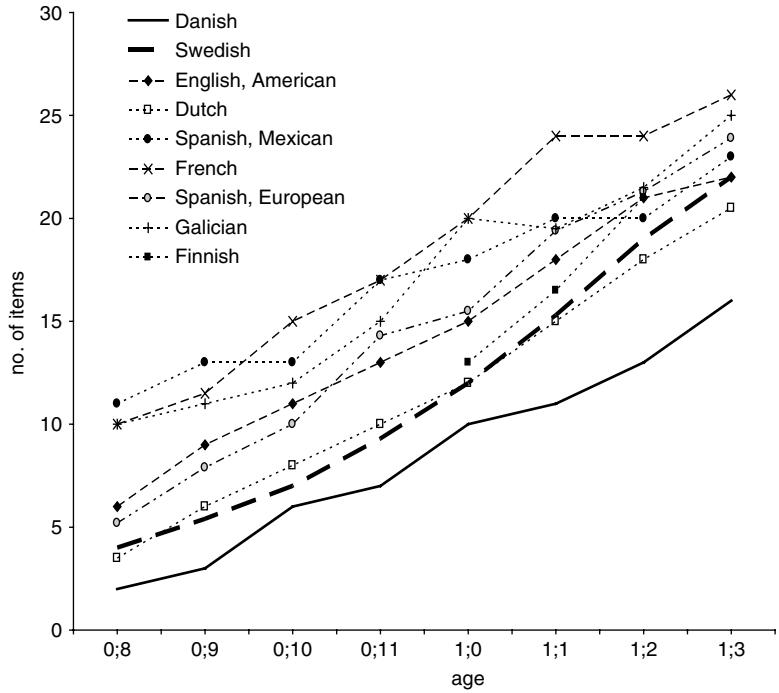


Fig. 3. Median comprehended phrases scores by age and study (*CDI: Words and Gestures*). NOTES: The Mandarin Chinese and the Italian studies only report mean values. The means for Mandarin Chinese are given in two age groups: 0;7–1;0: 9·4; 1;1–1;6: 15·0. The means for Italian are given in five age groups: 0;8–0;9: 6; 0;10–0;11: 10; 1;0–1;1: 16; 1;2–1;3: 17; 1;4–0;5: 20. The differences between the studies were significant ($p < 0{\cdot}0001$).

Is it possible to detect the same pattern in the early production of words? The median production score stratified by age and study (language) can be seen in Figure 4. The comparison comprises fourteen studies, including Danish.

The variation of the median score over all studies spans 0 words (several studies) to 4 words (Croatian) at age 0;8 to between 7 (Basque) and 27 words (Croatian) at age 1;3. Few children were reported to say any words in the earliest age groups (0;8–0;10) in any study. The level of scores is generally low in all studies also at age 1;3 (with Croatian and Galician children as marked exceptions), but nonetheless we can already observe consistent patterns in differences among the languages. Especially, the median scores of Danish children are among the lowest, but close to the Swedish children. As with comprehension, when comparing the variation of scores in the Danish study to the variation in the Swedish study, the Danish children score lower than the Swedish children at the 75th and 90th

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

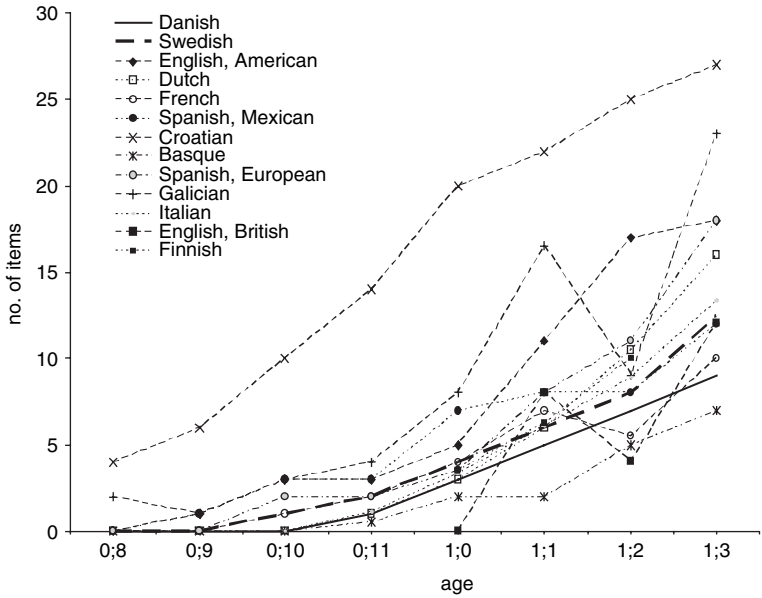


Fig. 4. Median vocabulary production scores by age and study (*CDI: Words and Gestures*). NOTES: The Chinese Mandarin study only reports mean values in one age group: 0;7–1;0: 0·9 (word). The differences between the studies were significant ($p < 0{\cdot}0001$).

percentiles at age 1;2 (75th percentiles: 12 words vs. 15·5 words in the Swedish data; 90th percentiles: 20 words vs. 27 words) and 1;3 (75th percentiles: 17 words vs. 25 words; 90th percentiles: 25 words vs. 46 words), whereas the other percentiles are nearly identical. The observed delay of the Danish children in the comprehension of words is not as marked for the early production of words even though the score of Danish children is among the lowest (together with the British English, French and Basque children).

A further section of *CDI: Words and Gestures* taps into processes associated with the early production of words by asking two questions related to the imitation and labelling of words (1C). In the first question, parents are asked to indicate how often the child imitates words; in the second question, how often the child names or labels things. Seven studies report the outcome of these measurements and the percentage of positives answers (i.e. 'always' and 'sometimes') to the two questions by age are shown for all studies in Figure 5 and Figure 6, respectively.

For Danish, the percentage of children who were reported to imitate or label shows a clear developmental trend. For imitation, the percentage ranges from 16·1% at age 0;8 to 59% at age 1;3, whereas labelling occurs

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

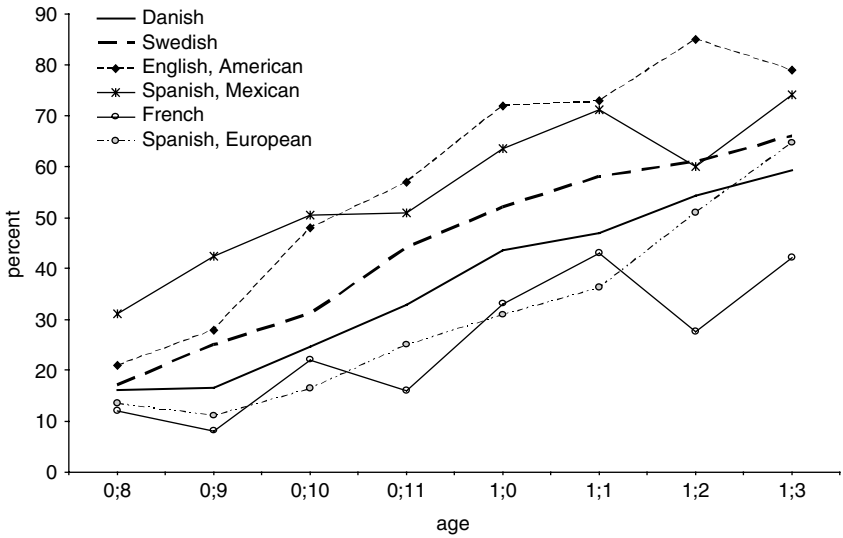


Fig. 5. Percentage of children who have started to imitate, by age and study (*CDI: Words and Gestures*).
NOTES: The percentage for Mandarin Chinese-speaking children, who are able to imitate, are given in an age group 1;1–1;6: 40%. The differences between the studies were significant ($p < 0{\cdot}0001$).

in only 2% of children at age 0;8 as compared to 38% at age 1;3. For imitation, the scores of the Danish children are in the middle/lower range compared to other studies, and for labelling, the Danish score is the second lowest. At each time point, Danish children imitate and label things to a lower extent than the Swedish children do.

From an overall point of view, the cross-linguistic comparison of the percentage of children reported to imitate and label suggests that the Danish children, parallel to the slower rate of word comprehension, also have a relatively slow start in building up their expressive vocabulary. Also in the production of words, as we saw in Figure 4, the median values of the Danish children were among the lowest scores found. What happens, then, if we expand the age range and consider the lexical development as measured in *CDI: Words and Sentences*. Can we determine a similar pattern here or do the Danish children catch up?

Figure 7 depicts the age profile of the median production score in the age range from 1;4 to 2;6. Sixteen studies (including the Danish study) are included in the comparison.

Figure 7 illustrates that the range of the median score is from 17 words (Swedish) to 103 words (European Spanish) at age 1;4 and between 380

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

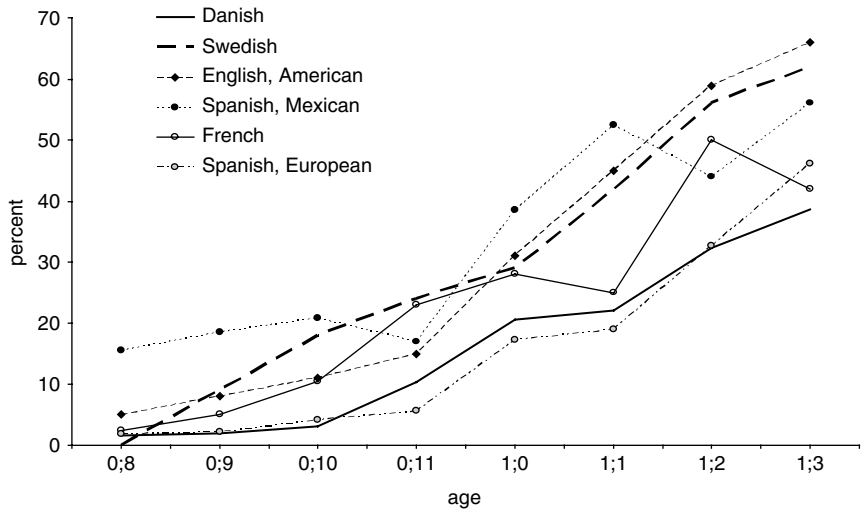


Fig. 6. Percentage of children who have started to label, by age and study (*CDI : Words and Gestures*).
NOTES: The percentage for Mandarin Chinese-speaking children, who are able to label, are given in an age group 1;1–1;6: 30%. The differences between the studies were significant ($p = 0{\cdot}0001$).

(French) and 561 words (Dutch) at age 2;6. Between 1;6 and 1;8 a sudden increase is apparent in all studies. When comparing median scores across studies we can observe that the Danish children start at the lower end, but begin to catch up around the age of 1;10 and finally end in the upper half. The scores of the Danish and Swedish children up to age 2;0 are almost identical, but from that age onwards the rate of acquisition is faster for the Danish children, resulting in a higher score up until age 2;4 (which is the oldest age group in the Swedish data).

DISCUSSION

In this paper, two main objectives have been put forward. The first and main objective was – through cross-linguistic comparisons – to outline the trajectory of Danish children's early lexical development relative to a number of other languages. To this end, we attempted to identify all studies which were methodologically comparable to the Danish study. There are, of course, differences in sample procedures and sample sizes across the studies, which may limit comparability across studies. One such limitation may be the differences in the number of words across the adaptations. In principle, it is somewhat of a drawback that the length of the word lists of the CDIs differs rather substantially across the different adaptations. This makes

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

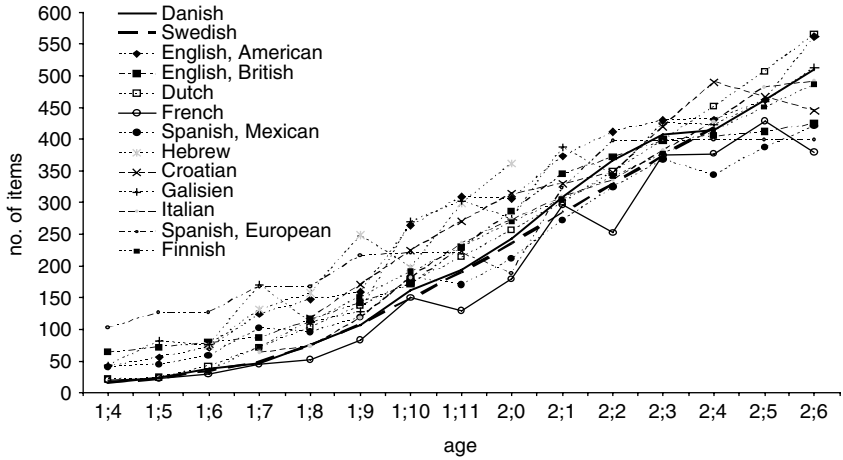


Fig. 7. Median vocabulary production score, by age and study (*CDI: Words and Sentences*). NOTES: The German, Icelandic and Mandarin Chinese studies only report mean values. The mean values for German (from the age of 1;6, listed as the first number), and Icelandic studies are given. The German and Icelandic results are read from graphs. The Danish mean values for each age group are indicated in parentheses as a comparison. 1;4: 58 (29); 1;5: 72 (41); 1;6: 65–134 (64); 1;7: 88–175 (75); 1;8: 104–157 (114); 1;9: 185–188 (134); 1;10: 214–254 (191); 1;11: 193–308 (206); 2;0: 285–303 (258); 2;1: 241–349 (304); 2;2: 259–344 (365); 2;3: 348–415 (390); 2;4: 392–411 (393); 2;5: 433–457 (432); 2;6: 421–503 (472). The mean from the Mandarin Chinese study is given in an age group: 1;7–2;0: 259·5. The differences between the studies were significant ($p < 0{\cdot}0001$).

some cross-linguistic comparisons difficult. However, the analyses presented in this paper, and especially those concerning our main hypothesis, mainly concern the very early vocabulary up to a few hundred words. As all adaptations of the CDI aim at catching the early vocabulary of children in the specific language/culture, one can assume that they in any case cover the very early vocabulary, covering the first 100–200 words. Lack of comparability may only be expected for analyses concerning larger vocabulary sizes.

Comparing the Danish results to these seventeen studies revealed several similarities, such as a large variation among children, acceleration in vocabulary growth in the second year of life and an asymmetry between vocabulary comprehension and production. The first conclusion from these comparisons, then, is that the developmental trend of Danish children's early lexical development is more or less equivalent to trends observed in the languages represented in the cross-linguistic study. From a broader perspective, this study adds to observations already acknowledged in the literature, i.e. children show strikingly similar patterns of development

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

in their early language acquisition across different languages (see Dale & Goodman, 2005).

Yet, when focusing attention on the actual size of the vocabulary at different age ranges, there are certain differences between the Danish and (almost) all of the other CDI-studies. The vocabulary comprehension score of the Danish children is the lowest across all studies from age 1;0 on (with the score in the British English study almost as low). The same tendency can be observed in relation to the early understanding of phrases. Contrasting the Danish study with the Scandinavian studies, Swedish children score higher than the Danish children, and with respect to the only available measure, also Icelandic children (the other Scandinavian language represented in the cross-linguistic study) score higher than the Danish children. Comparing the median comprehension scores between the Danish and Swedish data, it seems that the slower vocabulary growth rate in Danish does not affect the earliest phases (up to 60 words). This is supported by another observation: When taking the variability among children into account, it appears that the highest scoring Danish children (the 75th and the 90th percentile), i.e. children with a relatively large vocabulary, are scoring lower compared to the Swedish children, i.e. the delay affects the children with the largest vocabulary the most. From age 1;4, i.e. three to five months later compared to the majority of the studies, the growth curve for the Danish children becomes steeper, implying an acceleration in comprehension (cf. Bleses *et al.*, in press). Turning now to the vocabulary production score, up to age 1;3 Danish children are also among those with the lowest score, and the scores for imitation and labelling are likewise among the lowest. That Danish children are at the lower bound in the very early stage of production is not a surprising finding, as it is well known that there is an dissociation between the acquisition of the receptive and the productive vocabulary, with a 'fan effect' relating them (Dale & Goodman, 2005). Hence one should expect that a delay in acquiring the receptive vocabulary also imposes some delay in acquiring the productive vocabulary. Moreover, we have little reason to believe that the Danish sound structure can directly influence the acquisition of the productive vocabulary, hence we have to expect that the effect of the productive vocabulary vanishes over time. This is exactly what we find. After age 1;6, the acquisition rate increases, and at the end of the age range covered by the CDI-reports, the Danish children are in the upper half.

Before speculating about causal reasons for a delay in the comprehension of words in Danish children, we have to clarify whether the observed delay may just be a consequence of some type of reporting bias, i.e. whether Danish parents are more reluctant when filling out the parental report compared to parents in other countries. If this is the case, we should also expect under-reporting in other parts of the CDI. We have already seen

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

that we have no evidence of under-reporting on production at higher ages (*CDI: Words and Sentences*). Moreover, comparing the scores for gestures across studies, we find that the median total gesture scores in the Danish study by age is in the middle range when compared to the ten other studies which report scores for gestures (Swedish, American, Dutch, French, European Spanish, Galician, Italian, Mexican Spanish, Finnish and Chinese Mandarin), so we have neither evidence for under-reporting at the early age range where *CDI: Words and Gestures* are used, nor in the later age range where *CDI: Words and Sentences* are used (the Danish total gesture scores are reported in Bleses *et al.*, in press; see also Bleses *et al.*, 2007). By using similar lines of argument we can also argue against the view that differences in sampling techniques or other differences in the conduct of the studies included in our comparison can explain the observed differences in comprehension scores, as it is unlikely that any such difference affects only comprehension and acts only on comprehension in a specific time interval.

It is also worth mentioning, that – in spite of the tendency to lower comprehension scores – we find many similarities with respect to the interplay between different dimensions in the early language development when comparing the Danish study to other studies. First, we find that gestures (first communicative gestures), word comprehension and word production are all highly correlated at 12 months (Bleses *et al.*, 2007), a finding which has also been established for other languages (e.g. Almgren, Barnes, Colina & Garcia, 2005). Second, when examining the composition of the early lexicon in the longitudinal Danish CDI-study (cf. Wehberg, Vach, Bleses, Thomsen, Madsen & Basbøll, 2007), we do not find any differences with respect to the development of the composition of the early vocabulary in dependence on the vocabulary size relative to other studies published (cf. e.g. Maital *et al.*, 2000). Finally, we find significant (but rather small) gender differences in comprehension, which favours girls (Bleses *et al.*, in press, 2007) a finding also reported in some CDI-studies (cf. e.g. Fenson *et al.*, 2007).

In summary, these findings suggest that the lower comprehension scores are not due to the fact that Danish parents fill out forms differently, and that these lower scores are not associated with other fundamental differences in the structure of the early vocabulary development in Danish children compared to other languages, i.e. we have just a simple delay in the comprehension and very early production of words (and phrases). Why is it the case, then, that the Danish children apparently are somewhat delayed in the early comprehension of words?

We would like in particular to consider two possible explanations related to the hypotheses outlined in the Introduction. The first explanation may be the fact that the vast majority of Danish children are in daycare from

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

their first year of life, in contrast to parental care, which is still the more widespread situation in many countries. The (early) institutionalization may imply other socialization patterns, a different adult input or other differences compared to parental care, and this may affect the rate of language acquisition. However, if this were an explanation for the delay in comprehension, then we should also be able to demonstrate an influence of institutionalization on comprehension within our study. Indeed, we could observe a decrease in the comprehension scores from children who attend daycare a few hours per week (<25) to children who attend a medium number of hours per week (25–30 and 30–35) and children who attend daycare with a high number of hours per week (35–40 and >40), a pattern which was highly significant and consistently visible over the whole age range from 0;8 to 1;8 years. However, a direct comparison of children still at home with children attending daycare failed to produce a consistent difference, and the maximum observed difference in the median comprehension score was less than 12 words. Of course, all these comparisons may be hampered by the fact that (at least for children under 1;0) the parents' decisions regarding daycare are related to socio-economic factors and parental attitudes and the amount of hours in daycare per week may also influence the parents' opportunities to observe their child. In summary, we have strong evidence from our data that the effects of early institutionalization culture on comprehension in Danish are small, not least taking the rather large differences in proportion of children in formal daycare between Danish and other languages into consideration.

The second aspect we would like to consider is the possibility that the Danish sound structure has had an impact on the lower comprehension score. As indicated in the Introduction, Danish sound structure exhibits some very unusual characteristics. The vocalization of consonants and the frequent but not obligatory schwa-reductions result in long monotonous vocalic stretches, which make the Danish sound structure indistinct ('unclear') and weaken cues for word and syllable boundaries, or even make them outright non-existent.

Recently, an investigation was carried out studying the mutual intelligibility of the three mainland Scandinavian languages, Danish, Swedish and Norwegian (Delsing & Åkesson, 2005). The results of one of the comprehension tests, based on approximately 1,300 schoolchildren who were listening to news on the radio, showed that Danish is the Scandinavian language which is the most difficult to understand for other Scandinavians. Gooskens (2006) did some further analyses of this study in order to identify the predictors of intelligibility, including extralinguistic factors (contact and attitudes) as well as linguistic factors (phonetic similarity). Correlations revealed that phonetic distance was the main predictor of Danish being

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

most difficult to understand. This points to the importance of the Danish sound structure for the difficulties in comprehending Danish.

Do we have any indications that the difficulties experienced by the foreigners and second-language learners of Danish also cause the language acquiring Danish child any problems? First of all, we know that sound plays an important role for early language acquisition. To segment the sound chain into linguistic units is by no means a trivial task, due to the inherent characteristics of the speech signal. The characteristics of the speech signal make it complicated to segment and extract linguistic units, and the child therefore needs to identify relatively subtle cues to detect word boundaries and syllable boundaries. The more cues to word boundaries and syllable boundaries the input language offers, the easier it is for the child to identify linguistic units such as words in the input and the easier the task of identifying stable correspondences between sound and meaning for the child. We therefore assume that the fewer reliable cues to segmentation, the longer it takes to discover the correlations between the sound string and meaning and to obtain stable representations (see Bleses & Basbøll, 2004).

There is a large body of evidence documenting infants' use of different phonetic and prosodic cues in the input in order to segment and represent words, and a recent study by Kuhl and her colleagues also indicates that infants' early native vs. non-native discrimination skills at 7 months differentially predict future vocabulary development as measured by the CDI-method (Kuhl, Conboy, Padden, Nelson & Pruitt, 2005).

The developmental implications are that linguistic cues, documented in the literature as playing an important role in the early segmentation of words, are weakly signalled or in some cases entirely non-existent in Danish (see Bleses & Basbøll, 2004). Therefore we hypothesize that the Danish sound structure represents a difficult segmentation task not only for non-native speakers and second-language learners of Danish, but also for Danish language acquiring children, rendering some steps in early language acquisition so difficult – due to the scarceness of phonetic and prosodic cues – that it results in a delay in acquiring receptive abilities compared to children from other countries. The fact that the slower vocabulary growth rate in Danish children (cf. the comparison with Swedish children) does not affect the earliest phases of vocabulary development may support this hypothesis, as the very first words – like words for sounds, social expressions (like *hi* and *bye bye*) and social games, the child's own name, *mommy* and *daddy*, and *yes* and *no* to mention some of the most frequent – may probably not require the same segmentation skills as the later vocabulary because they have an extremely high frequency and (may) more often be said in isolation.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

Can we then infer that the Danish delay on vocabulary development observed in the data presented here is a result of children facing problems with the Danish sound structure, and that sound structure more generally affects children's early language acquisition abilities? The fact that children speaking British English score almost as low as the Danish children, whereas the American English-speaking children score considerably higher, apparently speaks against such a 'sound structure hypothesis'.

First of all, there is some methodological uncertainty related to the comparability of the result from the British English study to the other included CDI-studies due to the status of the adaptation by Hamilton *et al.* (2000) (recall that one of the inclusion criteria in our study was that the adaptation has to be as close as possible to the original American CDI). According to Hamilton *et al.* (2000), the CDI reports were initially used in Oxford as an aid to selecting materials for preferential looking studies. There have been three versions of the Oxford CDI, before the final fourth version was made, and there are substantial differences between the CDI word lists, both between the original American CDI and the Oxford CDIs, and within different versions of the Oxford CDI. Only the last version was intended to be similar to the American CDI. In the British English study, all the Oxford CDI data (based on the four different versions of the Oxford CDI) were amalgamated, and even though statistical analyses showed that no significant differences were found between the different versions of the CDI, we cannot discard the possibility that this has influenced the low score of the British English children.

However, if we accept that British English children are indeed slower than American children, can we then find any evidence of differences in the sound structure of the regional variants of English which may explain the different developmental patterns? Evidently, this question must be addressed directly, but there are at least two candidates which may facilitate the segmentation process of American children. First, and most important, final /r/ in words like *car* and *four* are pronounced, in contradiction to British English, thereby giving American children one extra segment and one extra cue for segmentation purposes (e.g. American English: *car* /kɑr/ and *four* /fɔr/; British English *car* /kɑ:/ and *four* /fɔ:/). The fact that final /r/ is not pronounced in British English causes British English to be more vowel-like word finally (as Danish), with possible consequences for the segmentation of the sound string into words. Second, in American English, at least in many words, there is a tendency to more secondary stresses than in British English (which has unstressed syllables instead), and thus a more salient prosodic pattern of words.

As stated above, the 'sound structure hypothesis' – and the postulated effect on segmentation between American and British English – would need to be explored explicitly. However, we regard it to be a fact that foreigners

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

and second-language learners experience difficulties in comprehending Danish (cf. Delsing & Åkesson, 2005), and that this supports such a hypothesis, but it is still necessary to address the question specifically in different ways in order to get closer to a more definitive answer. We believe that the CDI-data itself is a first, well-suited, source to investigate this issue further. This can be done by performing item-based analyses in the Danish data and by performing cross-linguistic comparisons on the basis of single items, especially those which are distinct in sound structure across languages, but are similar (or even identical) in terms of semantics and pragmatics. Not only the sound structure of the single item itself, of course, but also the sound structure of phrases typically surrounding an item, should (if possible) be taken into consideration. Combining CDI-data with knowledge of phonetic/phonological differences across languages will allow further tapping into the early comprehension of words on the basis of very large populations.

Finally, let us examine the second objective of the present paper, i.e. the appropriateness of the CDI-studies as a basis for cross-linguistic comparisons. First, the CDI-method – and all the CDI-adaptations and studies which have been conducted up until now – are a significant resource for cross-linguistic studies. The instrument is in the process of being adapted to more than 45 different languages. Out of the 32 completed adaptations, 28 can be regarded as 'standard adaptations', and are therefore highly comparable. So far data have been collected in more than 30 languages, including languages which differ extensively from a typological perspective, yielding a unique opportunity for performing close cross-linguistic comparisons on various components of early language acquisition. However, our process of compiling studies for the present paper indicates that it is unfortunately rather difficult to perform direct comparisons. In the first instance, this is due to difficulties in obtaining access to published material, not least manuals (which for some adaptations is the only place to obtain percentiles). Most studies are not published through regular channels, such as international journal articles or international book series. Second, it has been a surprise to us that less than half of all adaptations have been until now followed by a population-based study. Third, comparisons are complicated by the fact that the reported statistics differ so markedly across studies. A major task for future researchers using the CDI-instrument is therefore to attempt standardizing the reporting of studies in order to make comparisons more straightforward. Additionally, published data should be made available for other researchers.

The Center for Child Language and the CDI Advisory Board has developed a website for access by researchers and clinicians to CDI data in a comprehensive database called CLEX – Cross Linguistic Lexical Norm (see http://www.cdi-clex.org). This website will provide the opportunity for

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

researchers to explore full datasets (in the first place for the American, (Mexican) Spanish and Danish datasets), and allow users to make queries on almost all aspects of the dataset – from a single item, a single section of the parent report and up to total scores in the dataset – and make comparisons between the different languages where this makes sense. The aim is to include more languages, and thereby CLEX hopefully over time will provide a useful database for close cross-linguistic comparisons, giving rise to other new insights.

REFERENCES

Almgren, M., Barnes, J., Colina, A. & Garcia, S. I. (2005). Language acquisition in Basque children reflected by the MacArthur-Bates Communicative Development Inventories. Poster presented at the Xth International Congress for the Study of Child Language, IASCL, Berlin.

Atkinson, J., Anker, S., Braddick, O., Nokes, L., Mason, A. & Braddick, F. (2001). Visual and visiospatial development in young children with Williams syndrome. *Developmental Medicine & Child Neurology* **43**, 330–37.

Bleses, D. & Basbøll, H. (2004). The Danish sound structure – Implications for language acquisition in normal and hearing impaired populations. In E. Schmidt, U. Mikkelsen, I. Post, J. B. Simonsen & K. Fruensgaard (eds), *Brain, Hearing and Learning*. 20th Danavox Symposium, 2003, 165–90. Copenhagen: Holmen Center Tryk.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. & Basbøll, H. (in press). The Danish Communicative Development Inventories: validity and main developmental trends. *Journal of Child Language*.

Bleses, D., Vach, W., Wehberg, S., Faber, K. & Madsen, T. (2007). *Tidlig kommunikativ udvikling. Et værktøj til beskrivelse af sprogtilegnelse baseret på CDI-forældrerapportundersøgelser af danske normalthørende og hørehæmmede børn*. Odense: Syddansk Universitetsforlag.

Brink, L. (1979). Ordforrådets omfang og vækst. In M. Jansen & J. Lund (eds), *Børnenes sprog – sprogene omkring børn*, 107–175. Copenhagen: Hans Reitzels Forlag.

Caselli, C. M. & Casadio, P. (1995). *Il primo vocabolario del bambino: Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milan: FrancoAngeli.

Cepanec, M., Kuvac, J. & Kovačević, M. (2005). How much does KORALJE tell us about grammatical development of late talkers? Poster presented at the Xth International Congress for the Study of Child Language, IASCL, Berlin.

Dale, P. & Goodman, J. (2005). Commonality and individual differences in vocabulary growth. In M. Tomasello & D. I. Slobin (eds), *Beyond nature–nurture. Essays in honor of Elizabeth Bates*, 41–80. London: Lawrence Erlbaum Associates.

Delsing, L. O. & Åkesson, K. (2005). *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. Copenhagen: Nordiska ministerrådet.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J. & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language* **32**, 759–86.

Eriksson, M. & Berglund, E. (2002). Instruments, scoring manual and percentile levels of the Swedish Early Communicative Development Inventories, SECDI. *FoU-rapport* Nr. **17**. FoU-nämnden.

Fenson, L., Marchman, V. A., Thal, D., Dale, P., Reznick, S. & Bates, E. (2007). *The MacArthur Communicative Development Inventories: User's guide and technical manual*, 2nd ed. Baltimore, MD: Paul. H. Brookes Publishing Co.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

Fernández, M. C. & Umbel, V. M. (1991). *Inventario del Desarrollo de las Habilidades Comunicativas: Adaptación Cubana*. Miami, FL: University of Miami, Mailman Center for Child Development.

Fletcher, P., Chan, C., Wong, P., Stokes, S., Tardif, T. & Leung, S. (2004). The interface between phonetic and lexical abilities in early Cantonese language development. *Clinical Linguistics & Phonetics* **18**(6–8), 535–45.

Gooskens, C. (2006). Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. In J. van de Weijer & B. Los (eds), *Linguistics in the Netherlands* **23**, 101–113. Amsterdam: John Benjamins.

Grimm, H. & Doil, H. (2000). *Elternfragebogen fur die Früherkennung von Risikokindern. ELFRA-1: Elternfragebogen fur einjährige Kinder: Sprache, Gesten, Feinmotorik. ELFRA-2: Elternfragebogen fur zweijährige Kinder: Sprache und Kommunikation*. Göttingen: Hogrefe Verlag.

Grønnum, N. (2003). Why are the Danes so hard to understand? In H. Galberg Jacobsen, D. Bleses, T. O. Madsen & P. Thomsen (eds), *Take Danish – for instance: Linguistic Studies in honour of Hans Basbøll presented on the occasion of his 60th birthday 12 July 2003*, 119–30. Odense: University Press of Southern Denmark.

Hamadani, J. D., Hilaly, A., Mehreen, F., Yesmin, S., Tofail, F., Huda, S. N. & Grantham-McGregor, S. M. (2006). Does quality of psychosocial stimulation at home predict rural Bangladeshi children's development and behavior? Abstract presented at the 8th CAPGAN meeting, 6–8 February 2006, Dhaka, Bangladesh.

Hamilton, A., Plunkett, K. & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language* **27**, 689–705.

Hilaire, G., Kern, S., Viguié, A., Dudognon, P., Langue, J. & Romieu, J. (2001). Le développement communicatif des enfants français de 8 à 30 mois. *Le pédiatre*, **XXXVI**(182), 7–13.

Jackson-Maldonado, D., Thal, D., Fenson, L., Marchman, V., Newton, T. & Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual*. Baltimore, MD: Paul H. Brookes Publishing Co.

Jespersen, O. (1916). *Nutidssprog hos börn og voxne*. Copenhagen/Christiania: Gyldendal.

Kern, S. (2007). Lexicon development in French-speaking infants. *First Language*, **27**(2), 191–212.

Kim, M., McGregor, K. K. & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: language-general and language-specific patterns. *Journal of Child Language* **27**, 225–54.

Klee, T. & Stokes, S. (2005). Validity and preliminary norms for a British-English adaptation of the MacArthur-Bates CDI: Words and Sentences. Poster presented at the Xth International Congress for the Study of Child Language, IASCL, Berlin.

Kovačević, M., Jelaska, Z. & Brozovic, B. (1998). Comparing lexical and grammatical development in morphologically different languages. In A. Aksu-Koç, E. Erguvanli-Taylan, A. Sumru Özsoy & A. Küntag (eds), *Perspectives on language acquisition*, 368–83. Estambul: Bogaziçi University Printhouse.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T. & Pruitt, J. (2005). Early speech perception and later language development: Implications for the 'critical period'. *Language Learning and Development* **1**(3–4), 237–64.

López Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S. & Martínez, M. (2005). *Inventarios de Desarrollo Comunicativo MacArthur: Manual Técnico*. Madrid: Ediciones TEA.

Lytinen, P. (1999). Varhaisen kommunikaation ja kilen kehityksen arviointimenetekemä. Niilo Mäkii Instituutti. Jyväskylän Yliopiston Lapsitetkimuskeskus.

Maital, S. L., Dromi, E., Sagi, A. & Bornstein, M. H. (2000). The Hebrew Communicative Developmental Inventory: language specific properties and cross-linguistic generalizations. *Journal of Child Language* **27**, 43–67.

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press

Marschik, P. B., Einspieler, C., Garzarolli, B. & Prechtl, H. F. R. (2006). Events at early development: are they associated with early word production and neurodevelopmental abilities at the preschool age? *Early Human Development* **83**(2): 107–114.

Ogura, T., Yamashita, Y., Murase, T. & Dale, P. S. (1993). Some findings from the Japanese Early Communicative Development Inventories. *Memoirs of the Faculty of Education, Shimane University* **27**, 26–38.

Pérez Pereira, M., Resches, M. & Fernández, P. (2005). Interrelationships between components of language and communicative development in Galician as reflected by the MacArthur-Bates inventories. Poster presented at the Xth International Congress for the Study of Child Language, IASCL, Berlin.

Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language* **23**, 573–89.

Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language* **20**, 43–60.

Reese, E. & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language* **27**, 255–66.

Szagun, G. & Steinbrink, C. (2004). Typikalität und Variabilität in der frühkindlichen Sprachentwicklung: eine Studie mit einem Elternfragebogen. *Sprache Stimme Gehör* **28**, 137–45.

Teixeira, E. (2005). CDI's Adaptation to Brazilian Portuguese: Validation study of the words and sentences form. Unpublished paper presented at the Xth International Congress for the Study of Child Language, IASCL, Berlin.

Thordardottir, E. T. & Weismer, S. (1996). Language assessment via parent report: development of a screening instrument for Icelandic Children. *First language* **16**, 265–85.

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O. & Basbøll, H. (2007). Danish children's first words – analysing longitudinal data based on monthly CDI parental reports. *First Language*, **27**(4), 361–83.

Wu, J. (1997). Language, play and general development for Chinese infant-toddlers: Using adapted assessments. Unpublished doctoral dissertation, University of Colorado, Boulder.

Zink, I. & Lejaegere, M. (2002). N-CDIs: Lijsten voor Communicatieve Ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al. Acco, Leuven (Belgium)/Leusden (Netherlands).

https://doi.org/10.1017/S0305000908008714 Published online by Cambridge University Press