Beyond Philology No. 14/1, 2017
ISSN 1732-1220, eISSN 2451-1498

The use of the interjection *oh* across various age groups on the basis of the Abigail files in the CHILDES database

TOMASZ ŁAKOMSKI

Received 29.12.2016,
received in revised form 21.07.2017,
accepted 27.07.2017.

Abstract

The research undertaken aims to shed more light on the acquisition of the interjection *oh* on the basis of the Abigail files in the CHILDES database. The CHILDES database is a collection of transcripts of spoken interactions between the target child and his/her surroundings. The Abigail files comprise of data collected over four years, in which the child, Abigail was recorded at home at three-monthly intervals, a total of ten times. With regard to this paper, some approaches to interjections are sketched in the first part. Then, the general statistics of the use of *oh* are presented with reference to the functions it is used for by the participants of the interactions. Next, the most frequent three functions of *oh* are presented as calculated for the target child, the caregiver and a person of unknown age. Finally all functions of *oh* expressed in percentages with reference to the frequency of their occurrence are displayed.

Key words

discourse markers, interjections, language acquisition, language corpora

Używanie wykrzyknika *oh* przez różne grupy wiekowe na podstawie plików Abigail w bazie językowej CHILDES

Abstrakt

Artykuł stawia sobie za cel przeanalizowanie wykrzyknika *oh* na podstawie plików Abigail w bazie językowej CHILDES. Baza językowa CHILDES to zbiór transkryptów rozmów zapisanych w formacie CHAT. Pliki Abigail to seria rozmów dziecka z jego otoczeniem zarejestrowanych na przestrzeni czterech lat w trzymiesięcznych odstępach. W pierwszej kolejności artykuł prezentuje teoretyczne rozważania dotyczące wykrzykników. Kolejno przedstawiona zostanie ogólna statystyka użycia *oh* ze względu na funkcje, jakie pełni w wypowiedziach badanych uczestników interakcji. W następnej kolejności wskazane zostaną najczęstsze funkcje, jakie *oh* pełni w wypowiedziach badanego dziecka, jego opiekuna i osób trzecich. Na końcu przedstawione zostaną wszystkie funkcje, jakie *oh* pełni w wypowiedziach uczestników rozmów ze względu na ich częstotliwość.

Słowa kluczowe

korpusy językowe, przyswajanie języka, wykrzyknienia, znaczniki dyskursu

1. Interjections

There are few studies into the nature of interjections. They have been neglected in linguistic research due to their ambivalent nature as well as the difficulty to classify them according to the categories of description used in traditional grammar. Some researchers, however, treat them as discourse markers, which encompass a closed category of words usually reflecting the emotional state of mind and intentions of the speaker: for example, *well*, *ok, now*. In spite of this, the term discourse marker has not been used by all scholars to refer to the same group of lexical items. Lenk (1998), for instance, claims that two studies examining nearly the same occurrences may use

diverse terminology. The author makes references to Schourup (1985) who uses the term discourse particles, and Schiffrin (1987) who uses the term discourse markers when referring to lexical items which are largely identical. Blakemore (1987), meanwhile, invents the term discourse connectives. Fraser (2006), who uses the term pragmatic markers, mentions features which differentiate discourse markers from other parts of speech. According to this author, they are free morphemes usually occurring initially in the sentence, signalling a concrete message and being classified in terms of their semantic/pragmatic functions. Quirk at al. (1998) offers a more syntactic approach, referring to discourse markers as adjuncts, disjuncts and conjuncts depending on which position in the sentence they occupy.

Apart from formal classification, Aijmer (2013) describes how people process pragmatic markers. Thus, according to Aijmer (2013), speakers can constantly monitor and concurrently analyse what they are saying, and know how this corresponds to what others say in the interaction. Because of this, speakers are conversant with the nature of the interaction they participate in and can sense if, and when, the interaction is endangered. Therefore, pragmatic markers function as imprints of speakers' minds (Aijmer 2013: 4), reflecting their state of mind mental processes visible.

Interjections, a subclass of discourse markers, have been studied by scholars representing different theoretical approaches but they do not provide a unanimous definition of them. The first definition of interjections presented here was outlined by Wierzbicka:

> An interjection can be defined as a linguistic sign expressing the speaker's current mental state (1) which can be used on its own, (2) which expresses specifiable meaning, (3) which does not include other signs (with a specifiable meaning), (4) which is not homophonous with any other lexical item whose meaning would

> be included in its own meaning (that is, in the meaning of the putative interjection), and (5) which refers to the speaker's current mental state or mental act. (Wierzbicka 1992:164f.)

Another definition is offered by Ameka:

> Those words [...] are primary interjections that [...] are not used otherwise [...] Primary interjections are words or non-words which in terms of their distribution can constitute an utterance by themselves and do not normally enter into construction with other word classes, for example, *Ouch*!, ...*Oops*!, etc. They could be used as co-utterances with other units [...] Primary interjections tend to be phonologically and morphologically anomalous. (Ameka 1992a: 105)

Yet another definition of interjections has been given by Wilkins:

> A conventional lexical form which (commonly and) conventionally constitutes an utterance on its own, (typically) does not enter into construction with other word classes, is (usually) monomorphemic, and (generally) does not host inflectional or derivational morphemes. (Wilkins 1992: 124)

The definitions quoted above share certain features. According to all of them, interjections express the current mental states of their speakers, they usually stand on their own and do not normally come together with other word classes. Equally important, speakers utter them when they are on their own or when they are accompanied by other people in a social interaction. Additionally, they do not necessarily elicit any regular response from the interlocutor, although their absence may seriously impoverish the message conveyed by words. More importantly; however, their scarcity may lead to ambiguity and misunderstandings. Additionally, people displaying great control in spoken interactions, avoiding emotive reactions in response to what others say might be perceived as difficult to deal with.

In addition to the above, interjections are emotive reflections of the state of the mind of the speaker and, interestingly, they are universal, which means that can be traced in all languages. Quirk at al. (1998) treat interjections as closed units, that is a category which does not grow or permit much alteration to the existing forms. Other closed units are represented by articles, for instance. Biber, Leech, Conrad et al. (1999) define interjections as elements of emotive character illustrating the frame of mind of the speaker.

Stange (2009) claims that *oh* is the most frequent interjection of all. It normally appears at the beginning of a statement and is a response to what other people say. Apart from that, *oh* can also co-occur with other speech units. The most common combinations of *oh* are: *Oh yeah, Oh no, Oh well, Oh God, Oh I see, Oh right.*

Ameka (1992) divides interjections into primary and secondary types. Secondary interjections are differentiated by the fact that they mainly function as other parts of speech and are also used as interjections. Instances of secondary interjections are words such as *help!, fire!, careful!* These possess a basic meaning and are utilized interjectionally in a turbulent state of mind. They also usually demand some kind of reaction from other speakers.

Primary interjections are words which only have an interjectional function and apart from no independent meaning. Standard primary interjections comprise words such as *ouch!*, *wow!*, *gee!*, *oh!*, *oho!* Additionally, interjections are voiced as an overt response to verbal and non-verbal stimuli and can be construed only in the context in which they are used. Interjections construct a separate category because of their syntactical independence. Interjections can be used as discrete elements, and are always detached from the rest of the sentence (Ameka 1992). Moreover, they always construct an independent element of reference. Thus, they are not elements of the sentence in its entirety. This trait discerns interjections from other parts

of speech such as particles (Ameka 1992). Ameka (1992) suggests the following division of interjections.

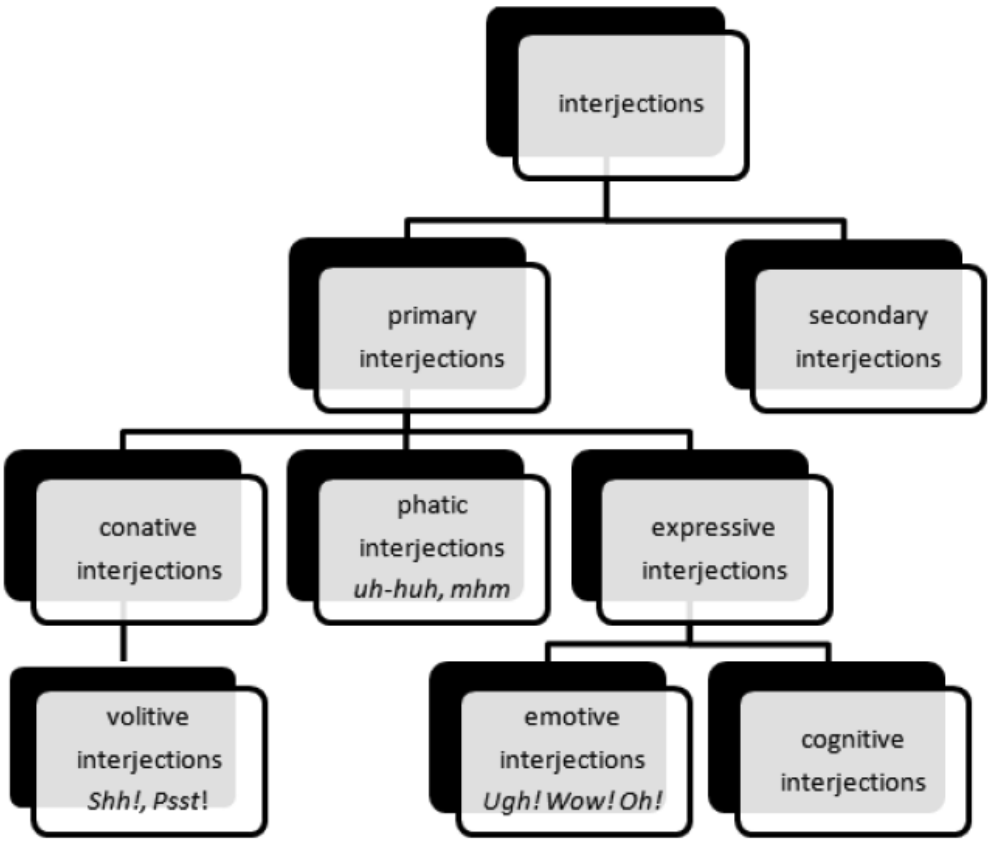

Figure 1
Ameka's (1992) classification of interjections

2. Aims of the study

In order to find out the role of *oh* in speech acts and the pragmatic functions it is used for, several research questions were outlined. They are presented below.

- When does the interjection *oh* first appear in the language repertoire of the children's speech analysed in the research?
- Which of the functions expressed by *oh* across the age groups (irritation, concern, lack of agreement, agreement, disappointment, praise, surprise, excitement, reported speech, attention getter, confirmation, disgust, something unpleasant, pain) are most frequent and which are least frequent?
- What are the general statistics of the examined interjection with respect to the recorded persons. Who (child, caregiver, person of unknown age) uses it most and who least frequently?
- What are the specific reasons for the production of *oh* across all the examined groups?

3. Database and method

3.1. Source and format of data

The database used to conduct the research for this study was CHILDES, i.e. Child Language Data Exchange System. The corpus was created in 1984 by Brian MacWhinney and Cathrine Snow as a tool to conduct research into first language acquisition. Today CHILDES consists of more than 130 corpora grouped according to contemporary languages. The transcripts of the researched files were in Codes for Human Analysis of Transcript (CHAT) format. Computerized Language Analysis was used to extract the data (CLAN). The codes are used to enable computers to read the transcript and perform search commands. CLAN is a statistical instrument used for calculating, among other things, the frequency or mean length of an utterance. The R programme was used to construct charts illustrating the results. R is a programming language

and software environment for statistical computing and graphics.

3.2. Data description

The Wells corpus was chosen for the following study because it illustrates adequately the targets that had been set for the study, i.e. it is an extensive corpus, with the samples recorded evenly over a long period of time. The whole Wells corpus consists of 299 files from 32 British children (16 girls and 16 boys) aged 1;6 to 5;0. The data was collected for four years, over which period each child was recorded at home at three-monthly intervals, a total of ten times. For this particular study ten files were chosen. These are all files that were registered with reference to one child, Abigail. For each recording session, the child was wearing a lightweight harness containing a radio microphone which transmitted continuously all the speech produced by the child and any speech by others, as well as noises that were loud enough for the child to hear. 24 examples of 90 seconds' duration at approximately 20-minute intervals between 9 a.m. and 6 p.m. were taken during each observation.[1]

3.3. Method

The CLAN programme was used to identify the files encompassing the selected interjection. The programme also extracted the tiers containing this interjection. In the CHILDES database tiers are divided into dependent and main tiers. Dependent tiers hold additional information, such as when the interaction takes place, who takes part in the interaction as well as details concerning the morphology and grammar of words forming strings of transcript. The main tiers are actual sequences of words uttered by participants of the dialogue. They

[1] Prepared on the basis of MacWhinney (2000).

are signalled and introduced by an asterisk. For example, *MOT means, that the words were uttered by the mother. Moreover, the searching string was set to extract 10 lines before and 5 lines after the interjection to ease its interpretation. All the instances of *oh* were taken into consideration, and these included the interjections produced by the child, the caregiver, usually the mother or a person of unknown age, usually represented by the child's sibling or the family's friend. More often than not, the CHAT transcript for the whole file had to be consulted in order to shed more light on the interpretation of the meaning of the interjection in a particular context.[2]

3.4. Results

After discarding those sets of lines which eluded interpretation the number of utterances which appeared in the database in the Abigail file was a total of 301 (127 utterances produced by children, 109 by adults, and 65 by persons of unknown age).

The earliest occurrence was at the age of 1;5.28 (year; month. day). The table above illustrates all of the appearances of *oh* in the overall section. The numbers of analysable utterances are shown, discarding those cases where the meaning of *oh* cannot be inferred clearly: for instance, this might be when the function of *oh* can be interpreted as belonging to more than one category. In such cases, the utterance was not taken into consideration.The table below visualizes all the functions of *oh* in the present study. The functions are enumerated, allocated to participants in view of the frequency of their occurrence. Finally the overall numbers of functions of *oh* against participants are given.

[2] Prepared on the basis of MacWhinney (2000).

Table 1
Results yielded for *Oh!* from the Wells corpus
for the Abigail file

	Overall	Adults	Children	People of unknown age
Total number	168	87	46	35
Analysable utterances	154	83	40	31

Table 2
Specific uses of *oh* across various speakers
taking part in the interaction

Participant \ Use	Concern	Praise	Surprise	Disappointment	Attention getter	Excitement	Something unpleasant	Irritation	Agreement	Lack of agreement	Confirmation	Reported speech	Disgust	Pain
Child	3	2	3	7	3	2	1	5	2	10	0	2	0	0
Care-giver	17	8	7	5	2	6	0	20	5	4	4	3	1	1
Un-known age	3	3	2	3	0	1	0	6	9	4	0	0	0	0
Total	23	13	12	15	5	9	1	31	16	18	4	5	1	1

Table 2 shows that the child uses *oh* mainly when she disagrees or when she is disappointed. The other conspicuous information emerging from the data is that the caregiver (mother or father) uses the interjection when they are irritated or con-

cerned. The person of unknown age uses the interjection when he or she agrees or is irritated. Overall, a general tendency surfaced: *Oh* was used most frequently when expressing irritation (20 per cent of all uses) and concern (15 per cent of all uses). Observations of mutual relationships between rearing children and their caregivers seem to confirm the collected data. Interactions of children with their surroundings, more often than not, are marked by a combination of irritation and concern on the part of parents. The other conspicuous information emerging from the data is the fact that in all three groups pain and disgust are not frequently expressed with *oh.*

Figure 2 illustrates the fact that *oh* is most frequently used by adults, followed by children and persons of unknown age. In the pie charts that follow, the main reasons for the use of *oh* registered for all three of the examined groups are given. At the present time, only three main functions have been established and the average value in percentages calculated, the results are shown in each of the pie charts presented in Figures 3, 4, 5 and 6.

Adults mainly use oh when they are irritated and concerned. This can be explained by the fact that parents are very often irritated when children, due to their lack of life experience, do not come up to their caregivers' expectations. For the same reason (deficiencies in knowledge of how to deal with obstacles) parents are concerned when children cannot overcome problems or cause themselves harm. The third function, praise, is equally explainable. When children do overcome problems mentioned earlier, parents are inclined to praise them with a supportive response.

Children mainly use *oh* when they disagree with somebody, when they are disappointed and irritated. It seems reasonable to expect that children, because of their inability to rightly perceive the intentions of the people surrounding them, will disagree with their caregivers. Having received a negative response, therefore, they are very likely to feel disappointed and irritated.

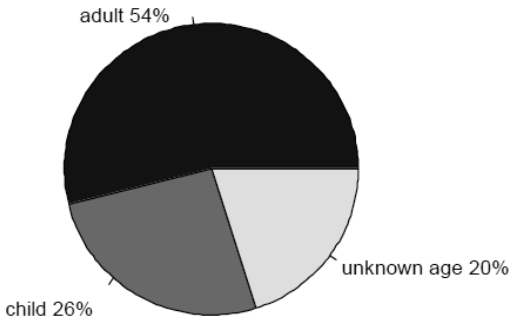

Figure 2
The overall use of *oh*

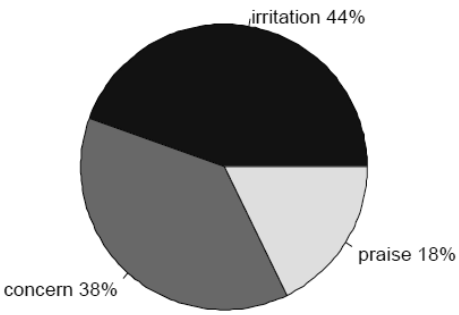

Figure 3
The three main functions of *oh* for adults

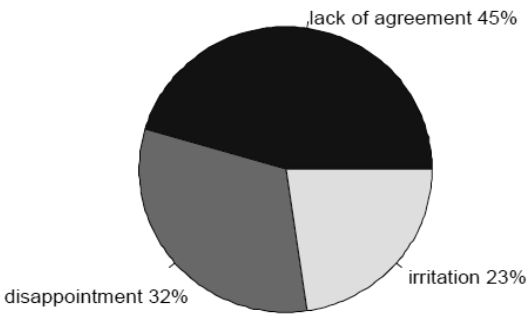

Figure 4
The three main functions of *oh* for the target child

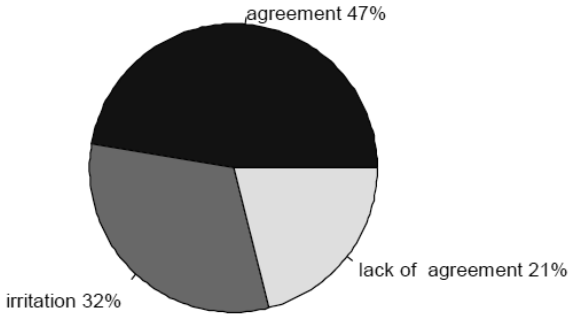

Figure 5
The three main functions of *oh* for the person of unknown age

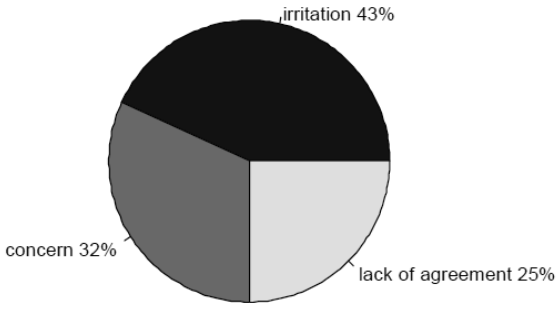

Figure 6
The three main functions of *oh* for all participants together

People of unknown age use *oh* when they agree, feel irritated or disagree. As already mentioned, the category of persons of unknown age comprises mainly of the child's siblings and family's friends. It has not been checked, but it seems reasonable to expect that the family's friends would most likely agree with the child in the presence of her parents.

The pie chart above illustrates the main functions *oh* performs in the examined files for all the participants taken together. It is assumed that the data is representative of the type of social interaction recorded. Since the central person of all the interactions was the child, it has an impact on the type of functions the interjection *oh* performs. In the research undertaken the main functions of *oh* might be summarized as a combination of irritation, concern and lack of agreement – intuitively and observationally confirmed by real life coexistence with small children.

The extracts from the CHILDES database below illustrate all the functions expressed with *oh* for all the participants taking part in the recordings. Fourteen functions of the interjection *oh* were selected for the study. These functions are presented here using the original transcript from the CHILDES database. The CHAT format uses symbols which are not inferable from the context without further explication. The exact length of the pauses between utterances in seconds are coded in the following manner: (10.). This means that there was a pause of ten seconds between the utterance and what followed. The speaker's identity is usually denoted by three letters. The code can be based either on the participant's name, as in *ABI or *REB, or on her role, as in *CHI or *MOT (child, mother). The symbol xxx is used when we cannot hear or infer what the speaker is saying. If there is a situation where a few unintelligible words cannot be recognized, several xxx strings may be used in a row. At times participants produce a wide diversity of sounds such as cries, sneezes and coughs. These are signalled in CHAT with the prefix &=, in order to produce forms such as &=sneezes and &=yells. Presumably the most prevailing is

&=laughs, which can be used to represent all types of laughs and chuckles. The symbols &ah, &hmm, &mm are used to indicate the diverse forms of the filled pauses. During the course of a conversation speakers often talk at the same time. The "overlap follows" symbol [>] denotes that the text encompassed by angle brackets is being produced at the same time as the following speaker's bracketed speech. It suggests that speakers are talking concurrently. This code is employed in connection with the "overlap precedes" [<] which means that the text coded in angle brackets is being uttered simultaneously with the preceding speaker's bracketed speech. Family-specific forms, illustrated by, for example breaky@ f[=breakfast], stand for child-invented speech that has been adopted by the whole family. Sometimes the origin of these forms are children, but they can also be older members of the family. In the example quoted, 'breaky' is used by the family and means 'breakfast'.[3]

What follows are extracts from the CHILDES database.

1. Disappointment. Here the mother expresses disappointment with the action undertaken by the child.

@Situation: Some toys fall down with a bang

*MOT: one at a time .

*MOT: pull them up .

*MOT: pull (4.) .

*MOT: come on (3.) .

*MOT: oh

[3] Prepared on the basis of MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

2. Concern. Here the target child expresses disappointment with her sister's action.

@Situation: M getting tea

*REB: oh yes I did .

*REB: oh yes I did .

*REB: oh no I didn't zzz (10.) .

*CHI: xxx .

*CHI: oh dear .

*CHI: a bang (2.)

3. Disagreement. The child Rebecca disagrees with the target child.

*MOT: &mm .

*REB: have bananas for lunch .

*REB: xxx we having xxx .

*REB: xxx we having xxx .

*REB: I want Ready_Brek .

*REB: oh no you can't .

4. Disappointment. Here the child Rebecca is disappointed.

REB: she she pee poo .

*MOT: xxx xxx [>] .

*REB: no [<] .

*REB: she's finished her (1.) breaky@f [= breakfast] .

*MOT: you haven't started yours have you ?

*REB: oh .

5. Irritation. Here the mother is irritated with Rebecca.

*MOT: xxx xxx [>] .
*REB: yes [<] .
*MOT: alright (3.) .
*REB: there yes (5.) .
*REB: baby .
*MOT: oh Becky just go off and do something else .

6. Praise. The family's friend praises Rebecca.

*FRE: grey ?
*FRE: yes .
*FRE: that's right .
*CHI: oh xxx (8.) .
*REB: can I get down please ?
*FRE: oh what a polite lady .

7. Surprise. Here the mother is surprised to find Abigail.

*MOT Abby (2)
*MOT oh there you are

8. Excitement. Mother is excited and laughs.

*CHI where is a duck
*MOT oh very good
*MOT &=laughs

9. Confirmation. The family's friend Erica confirms the information given by the mother.

*MOT conditioning and shampoo

*ERI oh I see

10. Something unpleasant. Mother expresses reaction to something unpleasant.

*HE I bang you with my spade

*CHI don't (4)

*MOT oh gosh

*MOT that was very bad

11. Pain. Here *oh* is used as a reaction to pain.

*MOT and your hand was there.

*MOT oh=ow[=ouch]

12. Reported speech. Here the child quotes somebody else's words

*CHI xxx(not recognizable)

*CHI she said oh no

13. Disgust. Here the mother expresses disgust.

@Time Start: 12:37

@Location: Car/Kitchen

@Activities: Involvement in nonplay

@Situation: they have just come home from the shops.

*MOT: oh yucky unloading .

14. Attention Getter. Here the child wants to have her mother's attention.

CHI:they're all swollen up .

*MOT: did you catch them under the piano when it fell (3.) ?

*MOT: oh golly .

*CHI: and it got softer (2.) .

*CHI: it got softer (18.) .

*CHI: oh look at this Australian card (3.) .

In Figure 7, all the fourteen functions are ascribed to all the participants of the recordings and represented using barplots.

The most frequent function of *oh* for the person of unknown age is agreement, and irritation for the adult. The least frequent functions for the groups mentioned are respectively pain and something unpleasant. The standard deviation for the person of unknown age is 2.722 and for the adult 5.837. The numbers suggest that the use of *oh* for adults is more varied than for persons of unknown age. The standard deviation for the person of unknown age is lower, thus its separate values are more concentrated around the arithmetic average. The use of *oh* for persons of unknown age is similar to that of the child.

The most frequent function of *oh* for the child is lack of agreement, and irritation for all the participants. The least frequent functions for the groups mentioned are respectively pain and pain. The standard deviation for the child is 2.824, and together 9.046. All the participants examined, when taken together with reference to functions of *oh,* display the greatest standard deviation. Its separate values are least concentrated around the arithmetic average. It suggests that functions of *oh* for persons recorded, when treated separately, are more alike than when summarized. The Abigail files, ten files of one child collected over a period of approximately four years

have demonstrated that the two main reasons for the production of *oh* are irritation and concern.

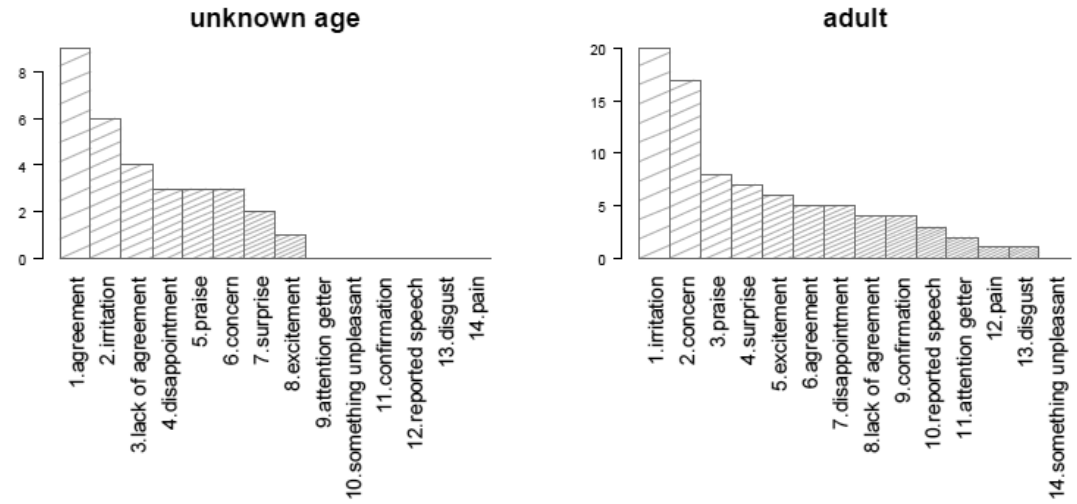

Figure 7
The specific functions of *oh* for adults and persons of unknown age

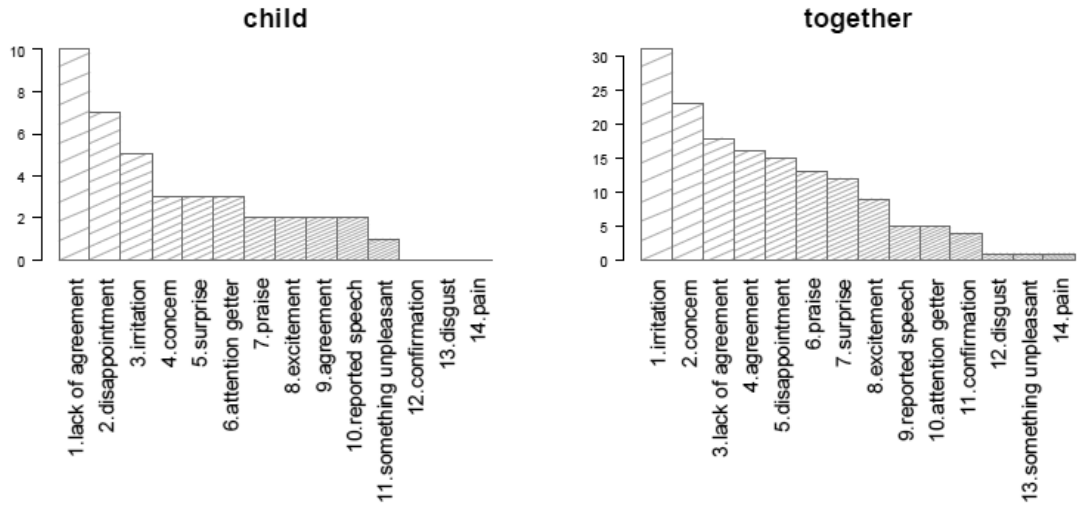

Figure 8
The specific functions of *oh* for the target child
and all the participants together

4. Discussion and conclusions

The earliest occurrence of the examined interjection was at the age of 1;5.28; however, it cannot be said if the child had already been producing it prior to the recording, i.e. at an even earlier age. Adults use the interjection in a more controlled manner and in more diverse contexts.

The main reasons for the production of *oh* in the examined files are irritation and concern. It seems reasonable to expect that the mutual relationship between children and their parents is full of concern on the part of the parents. This concern might lead to irritation of either the child or the parents if one side fails to satisfy the other. The adults in the researched files mainly use *oh* when they are irritated and concerned. It is generally acknowledged that parents care about their children and are concerned if something goes wrong. In addition, adults tend to control their feelings more and do not need to express disappointment, for example, every time they feel it.

As expected, the main reasons for the production of *oh* for children are disagreeing, being disappointed and irritated, which does not come as a surprise. Children, especially at an early age, test the world around them. They have certain preconceived notions about their surroundings and disagree if something runs counter to their will. This may also lead to disappointment and irritation. Other participants of the interaction mostly use the examined interjection when they agree, feel irritated and disagree. This group's uses of *oh* represent the fewest number of functions.

References

Aijmer, Karin (2013). *Understanding Pragmatic Markers*. Edinburgh: Edinburgh University Press.

Ameka, Felix (1992). "Interjections: the universal yet neglected part of speech". *Journal of Pragmatics* 18: 101-118.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited.

Blakemore, Diane (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Fraser, Bruce (2006). "Towards a theory of discourse markers". In: K. Fischer (ed.). *Approaches to Discourse Markers*. Oxford: Elsevier, 189-204.

Lenk, Uta (1998). "Discourse markers and global coherence in conversation". *Journal of Pragmatics* 30: 245-257.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Quirk, Randolph, Sidney Greenbaum (1998). *A University Grammar of English*. Harlow: Longman.

Schiffrin, Deborah (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Schourup, Lawrence (1985). *Common Discourse Particles in English Conversation*. New York: Garland.

Stange, Ulrike (2009). *The Acquisition of Interjection in Early Childhood*. Hamburg: Diplomica Verlag.

Wilkins, David (1992). "Interjections as deictis". *Journal of Pragmatics* 18: 119-158.

Wierzbicka, Anna (1992). "The semantics of interjections". *Journal of Pragmatics* 18: 159-192.

Tomasz Łakomski
Zespół Szkół Społecznych STO
im. Jana Nowaka-Jeziorańskiego
ul. Sobieskiego 7a
77-300 Człuchów
kierownikkolegiumchojnice@poczta.onet.pl