

## Deliverable D7: Demonstrator scenario

The aim of the COAVA project is to combine language acquisition with language variation research. The overarching research question is the following: is what children acquire very early, also less vulnerable to lexical variation?

In order to address this research question, let's consider the COAVA demonstrator scenario. First, open de COAVA tool:

<http://yago.meertens.knaw.nl/CoavaMainApplication/CoavaMainApplication.html>

The opening page shows all nouns that can be searched for and overlap maximally in both the childe database as in the dialect dictionaries for the Brabantian and Limburgian dialect area.

In the left corner, click on 'go immediately to'. There, one can find (i) the language acquisition tools (childe) and (ii) the language variation tools (dialect dictionaries). Ad (i) The childe search tool makes it possible to search for lexical items distributed according to the following independent variables (gender of the child, age of the child in intervals of three months between 15 months to older than 4 year), the name of the child (top 5) and the corpora involved (Schaerlaekens, Antwerp, Bol, Gillis, Groningen, vanKampen and Wijnen) and the date of the recordings. It is possible to search for 'en' and 'of' and 'literal sayings' or 'lexical nouns'. Ad (ii): the dialect lexicon search application shows the taxonomy nouns like 'het lichaam en de lichaamsdelen, karakter en gevoelens'. When searching for a noun, it shows the SIL entry, city/place and source, dialect area and lemma of the respective noun. The help function directs the user to the manual of the dialect dictionaries (sources, places, references).

Let's take a look at the openings page again. The first column presents the nouns in alphabetical order that are available both in the Childe and dialect dictionaries:

so immediately to...

.anguage Aquisition Tools ▶ Childes Search Browser  
.anguage Variation Tools ▶ Childes Charting Tool

the gap between language aquisition and

Childes Nouns	Childes	WBD	WLD
aai	<a href="#">3</a>	0	0
aap	<a href="#">47</a>	<a href="#">106</a>	<a href="#">171</a>
aardappel	<a href="#">26</a>	<a href="#">1076</a>	0
aardbei	<a href="#">4</a>	<a href="#">1112</a>	0
ajuin	<a href="#">2</a>	0	0
ananas	<a href="#">3</a>	0	0
anneke	<a href="#">1</a>	0	0
antoinette	<a href="#">2</a>	0	0
appel	<a href="#">40</a>	0	0
appelsien	<a href="#">31</a>	0	0
appelsiensap	<a href="#">2</a>	0	0
arm	<a href="#">3</a>	0	<a href="#">1</a>
armband	<a href="#">20</a>	<a href="#">565</a>	<a href="#">335</a>
arnold	<a href="#">448</a>	0	0
asperge	<a href="#">4</a>	<a href="#">44</a>	0
auto	<a href="#">453</a>	<a href="#">235</a>	<a href="#">184</a>
autobaan	<a href="#">3</a>	0	0
autobus	<a href="#">52</a>	<a href="#">31</a>	0
baan	<a href="#">2</a>	0	0
babbeke	<a href="#">1</a>	0	0

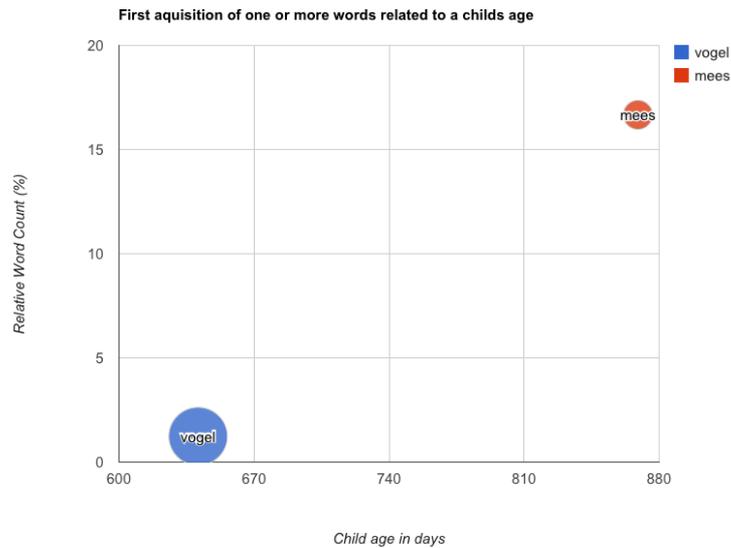
The second column in the page (presented) above presents the nouns that are uttered by many monolingual children speaking Dutch in the Netherlands and Belgium. These nouns are taken from the Childes/Talkbank project (see MacWhinney, B. 2007). The third column presents which nouns can be found in the Dictionary of the Brabantian dialects and the fourth column in the Dictionary of the Limburgian dialects. These three corpora have to be compared in order to address the research question: COVA gives us insight at what age children utter for the first time a certain noun (age of acquisition of a noun), whereas the two dialect dictionaries show us the amount of lexical variation of this specific noun. The amount of lexical variation is visualized in maps (see later).

The Childes charting tool (see below) provides which concepts have been acquired early 'vogel' for instance which is a basic level object or hyperonym and which ones have been acquired 'mees' for instance which is a hyponym and not a basic level object. This can be visualized by typing 'vogel' -> Language acquisition tools -> Childes charting tools -> Focus on first acquisition of a noun. The picture shows a blue circle visualizing that 'vogel' is acquired before 700 days, hence, before the year two. When typing 'mees', the graphic now reveals that 'mees' is much later acquired, i.e. just before 1000 days, thus roughly within 3 years. Childes charting tool is able to visualize both 'vogel' and 'mees' when typing both nouns (see below)

Please choose a view  
Focus on first acquisition of one or more nouns

Add words :  
    
 suggest terms/freq  
 quantity (max): 15

Select one or more corpora :  
 antwerp   
 ool   
 gillis   
 groningen   
 schaeerlaekens   
 vankampen   
 wijnen



With respect to 'vogel', the relative word count, presented in the chart above is between 1.0 and 1.5. This informs us about the correlation between word count and child age in days. The same information can be visualized in a bar diagram by clicking 'focus on a group of children':

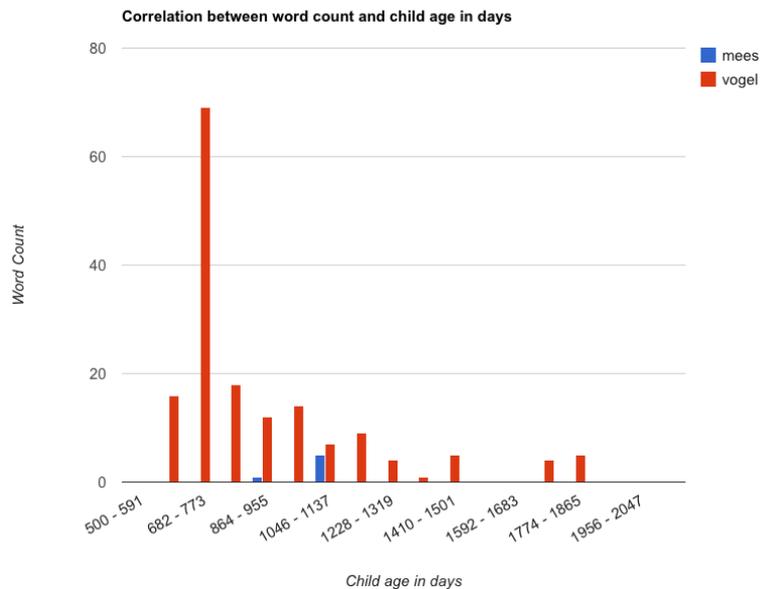
Please choose a view  
Focus on a group of children

Add words :  
    
 suggest terms/freq  
 quantity (max): 15

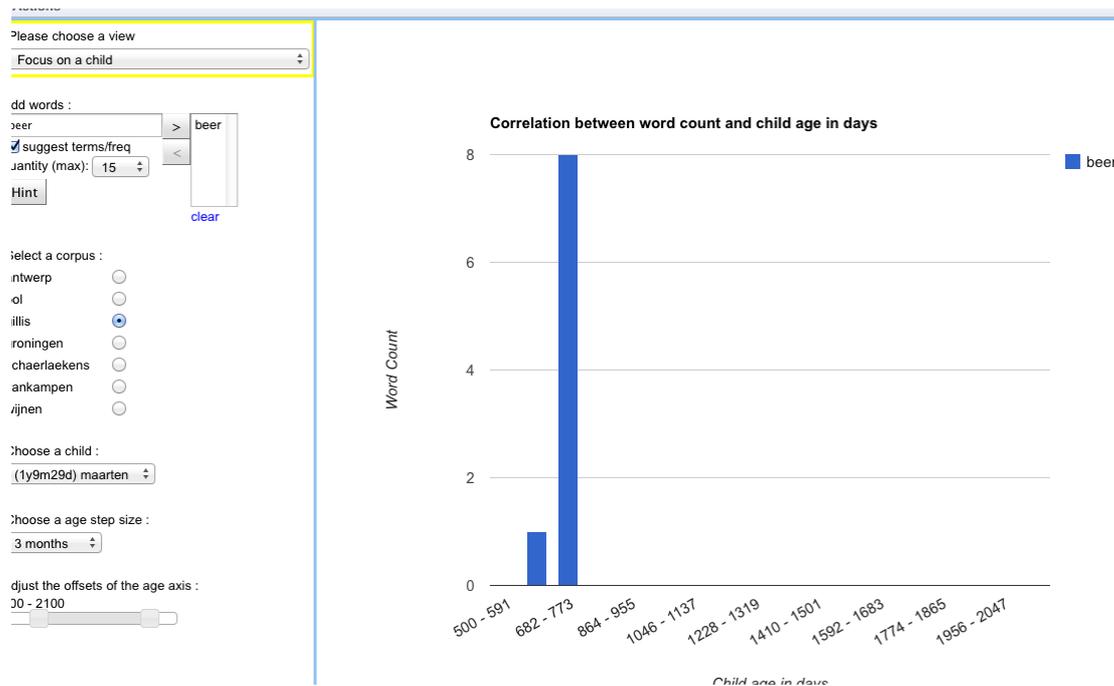
Select one or more corpora :  
 werp   
 s   
 ningen   
 aeerlaekens   
 ikampen   
 ten

Choose a age step size :

Just the offsets of the age axis :

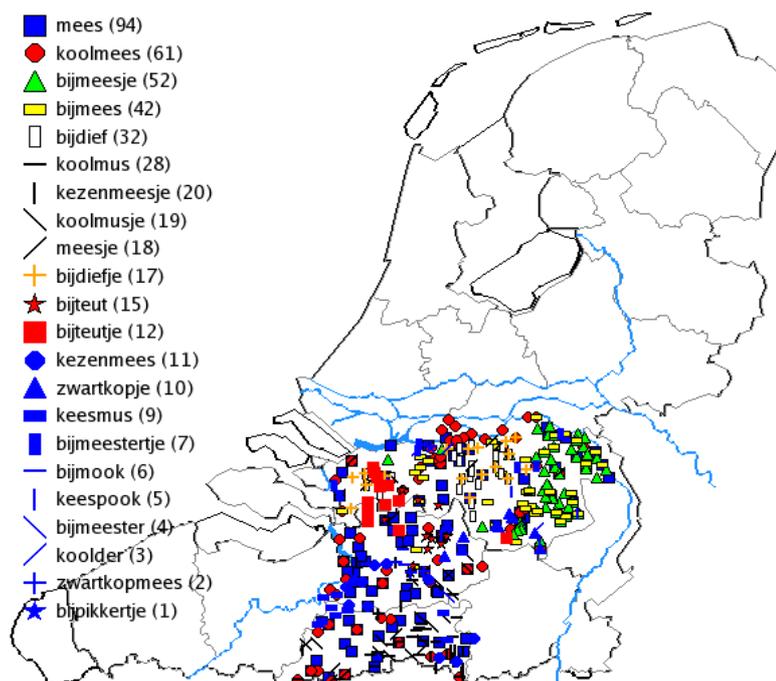


This tool also allows us to investigate the utterances of a specific noun at the level of the individual child for all recordings. The chart below shows us the word count of the noun 'beer' per three months of age of the child:

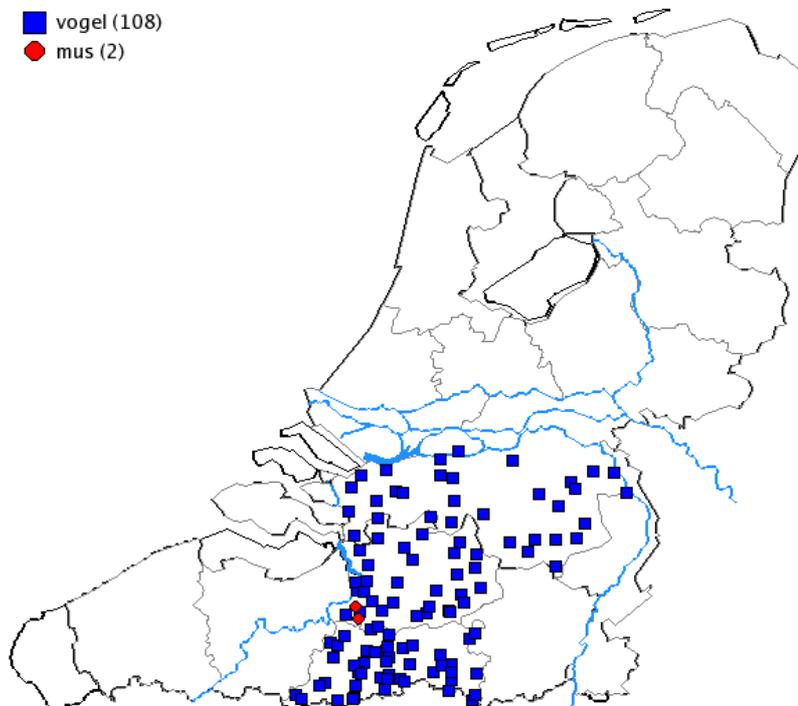


Until now, the tools visualized at what age children have acquired specific nouns i.e. their first utterances of the nouns. We are able to compare various nouns on word count and age of the child(ren).

Let us turn now the tools developed for language variation, in particular, dialect geography. For some concepts there is lot of lexical variation while for other concepts there is hardly any variation. Our research question implies that a hyperonym shows hardly lexical variation. By ticking 'mees' resp. 'vogel' in the Dutch Dialect Lexicon Search Application, an overview arises in which key words and SIL entries are displayed. The variation map shows the geographical distribution of dialect words for 'mees':



This stands in contrast to other concepts. Take for instance the denomination for ‘vogel’, that shows hardly any variation:



Basic level vocabulary (e.g. ‘vogel’) shows hardly any variation: (i) there is hardly lexical variation (only a blue square and red dot) and (ii) blue stretches out over a large geographical area. This is in contrast to hyponyms (e.g. ‘mees’), which show a huge number of different words for a concept, thus a massive amount of lexical variation. This concept may be regarded as less salient than its superordinate category, ‘vogel’.

We can compare lexical variation between various concepts, and thus measure variability, by calculating the number of lexical items per concept (Type Token Ratio). For ‘mees’ we found 104 types in 673 tokens: 104 different names in Brabantish dialects coming from 673 informants that filled in inquiry forms for one of the localities in this particular dialect area. In order to subdue the differences in numbers of informants in the lexicographical WBD-research we use Guiraud scores instead of simple TTR (Guiraud 1960). The Guiraud score  $G$  is 104 divided by the square root of 673 makes 4.01. The core component is able to perform an additional way of processing the results by computing measures. The COAVA tool as it is now, is able to compute quantitative measures indicating either the amount of geographic variation (in case of the dialect database) or acquisition age (in case of the acquisition database). In our Vienna talk (see COAVA website) we compared lexical learning (age of acquisition specific nouns) to various lexical variation measures (Type Token Ratio, Guiraud Scores, Silhouette index, Probability-Index).

The patterns in lexical variation and lexical learning indicate the measure of entrenchment of concepts. This tool allows researchers to confirm/reject the hypothesis mentioned above by posing questions.

