Journal of
Child Language Acquisition and Development
JCLAD

2016, September Vol 4 Issue 3 ISSN: 2148-1997

ISSN: 2148-1997

Table of Contents

Articles

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 78-97, 2016, September
ISSN: 2148-1997

Standard model of structures and contrast acquisition – proposal for phonological evaluation in the Brazilian Portuguese

Vanessa Giacchini[1]
Federal University of Santa Maria
Cristiane Lazzarotto-Volcão[2]
Federal University of Santa Catarina
Helena Bolli Mota[3]
Federal University of Santa Maria

Received : 25.10.2015
Accepted : 14.09.2016
Published : 30.09.2016

Abstract

Our aim was to spread out the analysis of the Standard Model of Contrast Acquisition to the syllable constituents of coda and complex onset, thus enabling an interaction between syllable structure and segmental inventory on phonological acquisition. We used analysis rating scales based on segmental strength and sonority, together with data from Brazilian Portuguese. The hypothesis for a Standard Model of Structures and Contrast Acquisition was created. This hypothesis suggests phonological acquisition of the simple onset happens in four stages, with stronger contrasts being acquired before the weaker ones. It also suggests coda acquisition happens in four stages, with final positions being acquired before middle positions, except when it comes to archiphoneme /R/, whose acquisition is co-occurrent. This hypothesis also suggests complex onset acquisition happens in two stages, beginning with the first consonant of complex onset and, on a further stage, command of the structure, regardless of segment class.

Keywords Acquisition, Speech, Language, Evaluation

1. Introduction

This new hypothesis for analysis is closely related to the ongoing development on the interaction between research on the process of phonological acquisition and the analysis of results based on phonological theories (Othero, 2005; Lazzarotto, 2007; Boersma, 2007; 2008; Matzenauer, 2012; 2013).

---

[1] Bio: Speech therapist, PhD in Human Communication Disorders and professor of speech pathology at the University of Passo Fundo, Rio Grande do Sul, Brazil. Her research interests are phonological acquisition, speech disorders, articulation, phonetics and acoustics early language acquisition. Corresponding author: fga.vanessa@hotmail.com

[2] Bio: Speech therapist, PhD in Applied Linguistics and a professor of linguistics at the Department of Language and Literature Vernacular Federal University of Santa Catarina, Santa Catarina, Brazil. Her research interests are the development of written language, speech disorder, phonetic and phonological acquisition. Corresponding author: cristiane.volcao@cce.ufsc.br

[3] Bio: Speech therapist, PhD in Applied Linguistics and associate professor of the Graduate Program in Human Communication Disorders of the Federal University of Santa Maria, Rio Grande do Sul, Brazil. Her research interests are segmental acquisition, phonology children, speech disorder and complex features. Corresponding author: nebolli@hotmail.com

Typical phonological acquisition happens when the child sets up a phonological system similar to the ones observed on their peers of the same age, i.e., matching the speech of their social group (Lamprecht, 1993; Matzenauer, 2012; 2013; Silva, Ferrante, Borsel & Pereira, 2012). In Brazilian Portuguese, this process happens gradually and in a non-linear way between birth and the age of 5:0, according to individual differences of each child (Lamprecht, 1993; 2004; Vitor & Cardoso-Martins, 2007; Toreti & Ribas, 2010; Silva, Ferrante, Borsel & Pereira, 2012).

However, this acquisition does not always happen as expected. Alterations to the regular process, referred as phonological disorders, might occur. These are usually defined as linguistic disorders manifested by using different patterns (Lazzarotto-Volcão, 2009; 2012; Henrich & Ribas, 2014). This disorder affects the phonological level of linguistic organization but not the mechanism of articulatory production (Grunwell, 1988; 1990).

Children with phonological disorders are usually diagnosed at an age above 5:0 and their phonological system is different from the input received from adults (Mota, 2001; Othero, 2005; Henrich & Ribas, 2014). The main feature of this group of subjects is their inability to systematically learn and organize their language sounds without having significant organic injuries affecting speech production, such as cognitive impairment, neuromuscular disorders, psychiatric disorders and environmental factors (Grunwell, 1988, 1990; Mota, 2001; Othero, 2005; Henrich & Ribas, 2014). Despite the non-existence of a noticeable organic problem, these children show an unexpected phonological system and they might express an incomplete segmental inventory when compared to their linguistic community standard (Mota, 2001; Wertzner, Pagan, Galea & Papp, 2007; Mezzomo, Giacchini, Dias, Luiz, & Lopes, 2011).

Phonological acquisition of a language is much more than just filling the segmental structure. It is imperative that syllable templates comprised by a language's phonology should be made available to the child (Miranda, 2012). Despite lack of consensus concerning the role of syllable structure in segmental acquisition, studies show a predisposition to acknowledge a strong connection between prosodic and segmental levels of phonology (Freitas, 1998; Galea & Wertzner, 2010; Miranda, 2012; Matzenauer, 2013; Mezzomo, Dias & Vargas, 2014).

According to the exposed by Selrirk (1982) the syllable is a structure composed by two constituents, the onset (O) and the rhyme (R), being the rhyme constituted by a sonority peak, the core (occupied by a vowel) and optionally the coda (Cd). The onset is not mandatory in the language and is liable of ramifications, the syllable structure can be observed on Figure 1. The syllable structure can be analyzed from three aspects: i) from the principles that determine the possible sequences of segments in the syllables; ii) the relation between the different syllabic constituents, and the segments that the languages license to occupy these constituents; iii) and of sonority, since the syllable is a sonority peak, and it is the one that conditions the factors "i" and "ii".

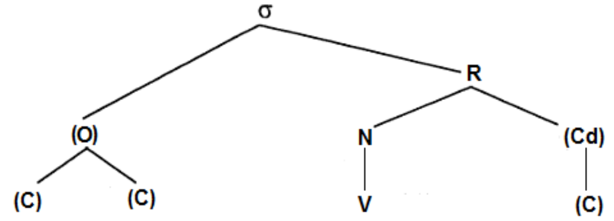


Figure 1. Representation of the syllable as ramified structure based on Selkirk (1982)

It is understood that the relations established for the proper and precise functioning of the linguistic systems, will also interfere in the phonological acquisition system, since it is necessary to observe and verify the relations established among the different phonological units during the acquisition process (Miranda, 2012; Matzenauer, 2013). The analysis of the interaction among these units (features, syllabic constituents, metric feet) allows the verification of how the phonological system of the child is being structured, and allows to verify if this process happens on a desired manner. The segmented analyses of only one of these units can generate mistaken conclusions about the acquisition process, verifying asymmetries that are not found when evaluating the different units in an interactive way (Matzenauer, 2015).

Acquisition data show a segment could be acquired as an onset but the same segment, on a given moment, could not be produced as a coda. This proves reliance between segmental and prosodic acquisition. On phonological disorders, the linguistic disorder affects both segmental and syllable level and reliance between segmental acquisition and syllable structure acquisition is noticeable (Freitas, 1998; Rose, 2000; Fikkert, 2007; Galea & Wertzner, 2010; Almeida, 2011; Yamaguchi, 2012; Lopes, Dias, & Mezzomo, 2015).

Using theoretical models to analyze speech data has provided a new interpretation and assessment of typical, as well as atypical, phonological acquisition. Research on clinical phonology has made possible a paradigm shift regarding assessment and therapy for children with atypical phonological acquisition. An example of such is the Standard Model of Contrast Acquisition (SCA) suggested by Lazzarotto-Volcão (2009), created from the hypothesis by Clements (2009). This model aims to describe phonological acquisition in Brazilian Portuguese via contrast acquisition in this language.

This model enables to evaluate a child´s sound system (not only with typical phonological acquisition, but also in the deviant cases) during its acquisition process and besides, it enables to assess the gravity of the deviant cases. The central idea of the SCA model is to enable to represent and explain the acquisition steps of the Brazilian Portuguese phonology, from the acquisition of the contrasts established by features and not on the acquisition of isolated segments. The model analyses the present, the absent and the in acquisition contrasts in the children´s system and the way they occupy the absent phonological spaces (Lazzarotto-Volcão, 2009).

As mentioned, the model proposed by Lazzarotto-Volcão (2009) for analysis, rating and therapy for phonological disorders was based on the phonological principles supported by Clements features (2009). This hypothesis used the Robustness Scale for consonant features to organize this model's stages for phonological acquisition. To this principle the authoress performed changes inherent to the Brazilian Portuguese, enabling to expose in a proper way the language acquisition process.

The Robustness Scale proposed by Clements (2009) was structured from the phonologies of the world language. From this, it is believed that the acquisition of a determined phonology can be guided with base on this principle. The Robustness Principle sustains that there is a universal features hierarchy, which reflects the preference of the languages in the constitution of their inventories. This preference is related to the position that the features occupy in a Robustness hierarchy. The features that occupy higher positions in the hierarchy are preferable to those which are in lower position in this ranking (Clements, 2009). Thus, it can be admitted in the development of the phonological system, the contrasts established by features ranked higher will be favored in their production, than those established by features ranked lower on this scale. And that the contrasts resulting from features that occupy the lower position on the scale, will be present only if the contrasts established by stronger features are already part of the child´s phonological system (Clements, 2009; Lazzarotto-Volcão, 2009). So the selection of this principle to validate stages for phonological acquisition expects the child to acquire stronger contrasts before acquiring weaker contrasts (Lazzarotto-Volcão, 2009).

Besides this principle, the authoress uses in the data analyses the principle of the Features Economy and the Marked Features Avoidance. The Principle of the Features Economy refers to the tendency of the traces to match maximally among each other (Clements, 2009). This principle can be observed through the voicing on the fricatives and plosives of the Brazilian Portuguese with only the features (+voice), the number of obstruent in the system is doubled. Thus, it is possible to verify that the feature is used with maximum efficiency on the system and on a very economical way by the language (Lazzarotto-Volcão, 2012). For Clements (2009) the Principle of Marked Features Avoidance can be seen as a new way for the students of the phonological inventories replacing the traditional way of analyses. The author considers the marking as a frequency effect which is determined by the appearance of the feature on the language. This way, the least frequent value of a feature is like this due to its unpredictability in the system, and, for this reason, can show special properties, many times associated with the marked values. Because of this, the author rejects the hypothesis of considering marked the feature which has more complex articulatory or perceptual properties, in comparison with other features that do not have them. On the basis of the marking idea Clements (2009) formalizes the Principle of Marked Features Avoidance in which a value of marked feature is considered when this is absent in some languages, on the contrary it is not marked. Having this principle as basis, the number of plosives must be bigger than the number of fricatives in a determined language.

From the phonological principles based on features of Clements (2009) and based empirically on the data typical phonological acquisition of Brazilian Portuguese, described in Lamprecht (2004), the Lazzarotto-Volcão (2009) structured the SAC model, which allows to represent and explain the stages of acquisition of phonology of Brazilian Portuguese, based on the acquisition of contrasts set the features and not the acquisition of isolated segments. A fundamental point of this model is the concept of formation of natural classes of segments from the action of features co-occurrence, as contrasts imply features co-occurrence and consequently natural classes of sounds in any phonological inventory or at any stage of the acquisition process (Lazzarotto-Volcão, 2009).

The SCA model (Lazzarotto-Volcão, 2009) suggests four acquisition stages, designed as periods when some language contrasts are stable in the children's system. These contrasts do not have a fixed order of appearance within these stages. These stages would be:
1st Acquisition Stage – Appearance and acquisition of contrasts, responsible for telling apart nasals from plosives, as well as command of point and sonority contrast in plosives context. It begins with the child's first productions and goes until 2:0 years. At this stage the features are marked acquired: [+sonorant] [labial], [dorsal], [-anterior], [+voiced]. By the end of this stage, the child has the following contrasts in their grammar: (i) sonorants versus obstruents; (ii) labials versus coronals in sonorant and non-sonorant contexts; (iii) labial-coronals versus dorsals, in obstruents context; (iv) sonorants versus non-sonorants, in plosives context. The child's system will have segments /p, b, t, d, k, g, m, n, ɲ/.
2nd Acquisition Stage – Command of contrast between non-continuant versus continuant, in obstruent context, allowing appearance of the fricatives class in the child's system. Acquisition of sonority and point contrasts in the same context also occurs. At the end of this stage the marked feature is acquired: [+continuant]. It begins around 1:8 years and goes approximately until 2:6 years. Segments /f, v, s, z/ are acquired in this stage.
3rd Acquisition Stage – Set up of contrast between nasals and orals in the sonorants class, due to appearance of the feature [±approximant]. Contrast between anterior and non-anterior in coronal fricative is acquired, as well as sonority contrast in the same context. At the end of this stage the marked feature is acquired: [+approximant]. Segments /l, ʃ, ʒ/ are acquired. This stage begins around 2:8 years and ends approximately at 3:0 years.
4th Acquisition Stage – Contrast between lateral approximant sonorants versus non lateral, due to the feature [±continuant] in liquids context. There is distinction of anterior versus non-anterior in laterals, as well as distinction of dorsal point versus coronal in non-laterals. At this stage there is no acquisition marked features. Segments /ʎ, R, ɾ/ are acquired. This stage begins around 3:4 years and goes until 4:2 years.
By using phonological models for clinical description and speech data analysis, this study aims to add analysis and description of coda and complex onset syllable structures to the hypothesis by Lazzarotto-Volcão (2009). This will allow adding the interaction between syllable structure and segmental inventory on phonological acquisition, as there are very few assessment hypotheses containing syllable-segment interface.

2. Methodology and results

2.1. Procedures for the organization of the syllabic structure analysis proposal on the Standard Model of Contrasts Acquisition

In order to extend Lazzarotto-Volcão model (2009) by adding analysis of coda and complex onset syllable structures, specific diagrams for each syllable constituent were created.
Firstly, an acquisition hypothesis was drafted, based on the Robustness Scale for Co-occurrence of Consonant Features (Table 1) (Lazzarotto-Volcão, 2009; 2010). This scale shows the co-occurrences that are established

through the acquisition. Being the highest placed co-occurrences in the hierarchy the most robust and those arranged below the less robust.

Table 1
Robustness Scale for Consonants Features of Co-occurrences for the Acquisition of Brazilian Portuguese, by Lazzarotto-Volcão (2009)

Groups	**Co-occurrences for Consonants Features**
Group a	[± sonorant] [-sonorant, -continuant, coronal] [-sonorant, -continuant, labial] [-sonorant, -continuant, dorsal] [-sonorant, -continuant, ±voiced] [+sonorant, -aproximant, labial] [+sonorant, -aproximant, coronal] [+sonorant, -aproximant, coronal, ±anterior]
Group b	[-sonorant, ± continuant] [-sonorant, +continuant, coronal] [-sonorant, +continuant, labial] [-sonorant, +continuant, coronal, ±voiced] [-sonorant, +continuant, labial, ±voiced]
Group c	[-sonorant, +continuant, coronal, ±anterior] [-sonorant, +continuant, coronal, -anterior, ±voiced] [+sonorant, ±approximant]
Group d	[+sonorant, +approximant, ±continuant] [+sonorant, +approximant, -continuant, ±anterior] [+sonorant, +approximant, +continuant, coronal] [+sonorant, +approximant, +continuant, dorsal]

This hypothesis (based in Robustness Scale for Co-occurrence of Consonant Features) assumed that segments in several syllable positions would be acquired according to the rating scale. The child would thus acquire stronger contrasts in a stage previous to acquisition of weaker contrasts. Afterwards, this hypothesis was analysed with empirical data from Brazilian Portuguese available at Lamprecht (2004) and thus the hypothesis of integrating syllable structure in the SCA model was validated.

Regarding segments used in the preparation of this hypothesis, segments /N/, /S/, /L/ and /R/ were taken into account, as these act as a decreasing margin in a syllable in Brazilian Portuguese (Wetzels, 1997; 2000; Freitas, 1998; Lamprecht, 2004; Galea, & Wertzner, 2010; Mezzomo, Dias, & Vargas, 2014).

In the complex onset position, the following obstruents were admitted in C^1 position: /p, b, t, d, k, g, f, v/; and liquids /l, ɾ/ in C^2 position (Lamprecht, 2004). However, not all of these combinations are possible nor found in Brazilian Portuguese. For example, [vr] only appears in the middle; [vl] only appears on loan words from other languages; and [tl] seldom occurs and only in the middle preceded by the vowel 'a' (Freitas, 1998; Ribas, 2003; Lamprecht, 2004; Montenegro, 2012).

For determining the status of contrast acquisition, we adopted the proportions suggested by Yavas, Hernandorena & Lamprecht (2001). In these, contrast or syllable structure was deemed absent when proportion of correct production was equal to or less than 50% in each assessed context; unstable, when proportion of correct production was between 51% a 75% in the assessed context; and acquired, when proportion of correct production was above 76% in each context.

Specific notations, colors and shapes were used for each acquisition stage, for each syllable constituents and for acquisition status. In acquisition stages, we used red for the first stage; blue for the second; yellow for the third; and green for the last acquisition stage of contrasts or structures (Table 2).

For setting syllable constituents apart, we used different shapes: the square shows simple onset; the rectangle middle coda; the oval shape, end coda structure; and the conjoined rectangles, complex onset (the first rectangle shows the obstruent and the second shows the liquid) (Table 2).

As for acquisition status, different fillings were used. When contrasts, co-occurrences and syllable structures are absent, there is no filling in their specific shapes; if contrast or syllable structure is being acquired, the filling shows diagonal lines; when contrasts, co-occurrences and syllable structures have been acquired, the filling is solid (Table 2).

Table 2
Notes concerning the stage of acquisition, the evaluated structures and the status of acquisition of contrasts or syllabic structures used in the expansion of the model

Colors	**Shapes**	**Acquisition status**
→ 1st stage - simple onset and coda	→ Simple onset	→ Contrasts, co-occurrences and syllable structures are absent
→ 2nd stage - simple onset and coda	→ Middle coda	→ Contrasts or syllable structures in acquisition
→ 3rd stage - simple onset and coda	→ End coda	/x/ → Contrasts, co-occurrences and syllable structures in acquisition
→ 4th stage - simple onset and coda	C¹ C² → Complex onset	→ Contrasts or syllable structures acquired
→ 1st stage - complex onset		/x/ → Contrasts, co-occurrences and syllable structures acquired
→ 2nd stage - complex onset		
→ Availability contrast or the structure		

2.2. *Proposal of the Acquisition Standard Model of Contrasts and Structures (SCA-E)*

Diagrams were created for each analysed syllable structure. For simple onset structure, we kept the hypothesis presented by Lazzarotto-Volcão (2009), with a few structural modifications. For coda and complex onset constituents, we created diagrams based on the Robustness Scale for Co-occurrence of Consonant Features (Lazzarotto-Volcão, 2009) and the Sonority Rating Scale by Clements and Hume (1995).

It is known that acquisition happens in a joint and interconnected way, so the suggested separation was merely for didactic and organizational purposes. In this hypothesis, we kept the same features exposed in the original model while respecting the theoretical basis matching the analysis of syllable structures. The following will be presented each of the diagrams drawn to the syllabic constituents, simple onset, coda and complex onset.

2.2.1. *Proposal for a Standard Model of Structures and Contrast Acquisition – Simple Onset*

On the simple onset syllabic constituent analysis proposed, the same idea presented on the original SCA model was kept (Lazzarotto-Volcão, 2009). The syllabic structure of the simple onset is formed by only one consonant, and it is considered the least restrictive position on the Brazilian Portuguese, since all the consonants of the language can occupy such position. Emphasizing that the phonemes /ʎ/, /ɲ/ and /ɾ/ happen in syllables located in the middle of the word.

To the diagram elaborated by Lazzarotto-Volcão (2009) two small alterations of structural character were performed, which do not interfere on the content and on the reading of the model. The changes are: i) addition of a rectangle on the top of the picture with the indication of simple onset structure; ii) insertion of overwritten numbers on the nasal, liquids /l/ and /ɾ/ and on the deaf coronal fricative /s/.

The i) modification was needed given the design of two other diagrams, one for coda and the other for complex onset. This way, each diagram has the rectangle concerning the represented syllable structure. In addition, every time this structure is available in the child's system, the rectangle will be filled with grey, showing the child already commands that structure.

The ii) modification was added to the diagram because nasals, liquids /l/ and /ɾ/ and non-sonorant coronal fricative /s/ can be both in simple onset position and in coda position. Thus, so that acquisition order for those constituents could be shown in the system, we added the number – 1 or 2 – inside the segment specific rectangle. When observed on the nasals, liquids /l/ and /ɾ/ and on the deaf coronal fricative /s/, the overwritten number 1 indicates that this segment was acquired beforehand on this syllabic constituent, than this same segment that contains the overwritten number 2. When these numbers appear, one can determine on which syllable position the segment was first acquired. With the suggested modifications, the hypothesis for analyzing and describing data on simple onset for Brazilian Portuguese is shown in Figure 2.

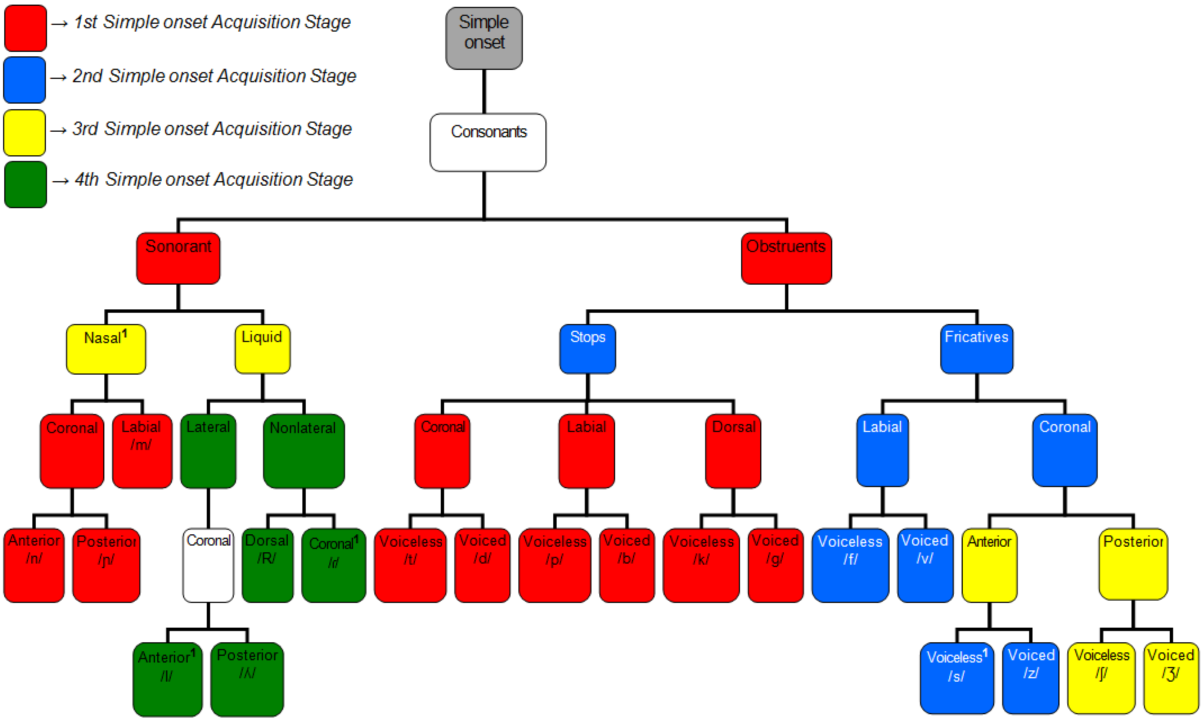


Figure 2. Standard Model of Structures and Contrast Acquisition for Simple Onset

Based on the model it is possible to represent and explain the stages of acquisition of phonology of Brazilian Portuguese, by acquiring contrasts established by the features and not the acquisition of isolated segments. Thus, it is possible to verify the formation of natural classes segments from the action of co-occurrences of features, whereas contrasts require the co-occurrences of features and hence the existence of natural classes of sounds in any phonological inventory or any step of acquisition process.

2.2.2. Proposal for a Standard Model of Structures and Contrast Acquisition – Middle Coda and End Coda

The coda syllabic structure can be filled by one or two phonemes that occupy the right margin of the syllable. On the Brazilian Portuguese this constituent presents some restrictions regarding to the segmental material that can occupy such position. The language allows the coda to be filled by a nasal sonorant, with no specifications for articulation point features, traditionally a consonant nasal archiphoneme, a coronal lateral sonorant or by a coronal continuous obstruent, traditionally represented by the archiphoneme /S/ (Câmara Jr. 1973).

Using empiric data of the Brazilian Portuguese with the proposal based on the Robustness Scale for co-occurrence of consonant features (Lazzarotto-Volcão, 2009), a sequence of acquisition of the constituent was proposed. Having as basis the data available in Lamprecht (2004) analyzed with the co-occurrences previewed on the Robustness Scale by Lazzarotto-Volcão (2009) the coda acquisition would happen in four steps, which would be:

1st Coda Acquisition Stage (red) – Acquisition of archiphonemes /N/ and /L/ in end position. This stage begins around 1:4 years and goes until approximately 1:7 years. In this stage, end coda is performed adequately by archiphonemes /N/ and /L/. Comparing to the Robustness Scale for Co-occurrence of Consonant Features (Lazzarotto-Volcão, 2009), an agreement about early acquisition of the nasals is verified. Studies indicate that the acquisition of the nasal placed at the end of the word happen at very initial periods of the normal phonological development, since the omission of this segment does not manifest itself anymore at two years of age (Mezzomo, 2005; Mezzomo, Quintas, Savoldi, Bruno, 2010).

The precocity of the final coda acquisition with the lateral liquid diverges from the observed when this segment is in position of simple onset. This disagreement can be due to the vocalic nature that this liquid acquires in coda position in the Brazilian Portuguese speakers (ex. 'mel' → ['mɛw]).

As the vocalic system is acquired quite early by children and the lateral liquid is similar to the vowel [u] during phonetic production, this could explain early acquisition of the segment in end coda position (Lamprecht, 1993; Mezzomo, 2005; Lamprecht, 2004; Mezzomo, Quintas, Savoldi, & Bruno, 2010; Galea & Wertzner, 2010, Mezzomo, Dias, & Vargas, 2014). On the European Portuguese (Lousada, Mendes, Valente & Hall 2012; Amorim, 2014; 2015) this behavior is not observed and the coda with the lateral liquid tends to be acquired before the middle position, differing from the observed on the Brazilian children.

2nd Coda Acquisition Stage (blue) – Acquisition of nasal archiphoneme in middle position and of fricative archiphoneme /S/ in end position. Thus, both sonorants and obstruents can fill this position. This stage begins around 1:6 and goes until 2:6 years. It completes acquisition of archiphoneme /N/ in both coda positions.

According to the hypothesis based on the Robustness Scale for Co-occurrence of Consonant Features, fricative acquisition would happen after acquiring nasals and before acquiring liquids and both positions would be acquired simultaneously, which was not observed. The constituent acquisition period was affected by coda position in the word. The role of syllable position and the role of the word affect greatly the acquisition process of coda constituents.

In the second expected acquisition stage of the coda constituent, the system already has three codas in end position - /N/, /L/ and /S/ - and just one in middle position, /N/. End codas seem to benefit from the acquisition process. According to Freitas, Miguel & Faria (2001), word end position has a major role in processing morphological information, which could explain the preference for end codas before middle codas. Its acquisition benefits from its more prominent morphological position and this morphological piece of information usually transported by the final coda would be more relevant for the acquisition of the fricatives in coda position. It can be inferred that the children adopt different acquisition routes during the construction of the architecture of the distinctive features according to the syllabic structure that the segment occupies (Almeida, Costa, Freitas, 2010).

This tendency for end coda position to be stable before middle position has been shown in other studies in this field. These studies show phonemes in end position are privileged over middle position phonemes (Lamprecht 1993; Mezzomo, 2005; Lamprecht, 2004; Mezzomo, Quintas, Savoldi, & Bruno, 2010; Galea & Wertzner, 2010; Mezzomo, et al, 2011; Miranda, 2012; Mezzomo, Dias, & Vargas, 2014).

3rd Coda Acquisition Stage (yellow) – It predicts middle coda acquisition, such as /S/ e /L/. It begins around 2:6 and goes until 3:0 years. By the end of this stage, the system has codas with /N/, /S/ and /L/ in both positions. Only command of coda with archiphoneme /R/ is missing. Once more, coda position in the word affects the segments acquisition process, which shows the relevance of coda structure position for that constituent's acquisition. Mezzomo, Quintas, Savoldi, & Bruno (2010) highlight coda segments are firstly acquired in end position, no matter the sound class. The authors also point out the role of tone for favouring the perceptual edge in a syllable margin.

A study performed with a monolingual child (European Portuguese) and a bilingual child (European Portuguese and French) verified that the two children present a different route on the acquisition of fricatives on the onset and coda position. This variation can be seen as a clue that the children build a feature architecture according to the syllabic structure where the fricative is made (Almeida, Costa & Freitas, 2010). The influence of the syllable role on the emergency of the segments was also observed on studies with the Brazilian Portuguese (Matzenauer, 2013).

4th Coda Acquisition Stage (green) – Predicted acquisition of archiphoneme /R/ in middle and end coda. This stage begins around 2:10 and goes until 3:10 years, marking the end of coda constituent acquisition in Brazilian Portuguese. This was the only stage matching what was suggested by the hypothesis supported by the Robustness Scale for Co-occurrence of Consonant Features. As predicted, coda formed with archiphoneme /R/ was the last to be acquired and its acquisition was co-occurrent in both middle and end positions. According to data from Lamprecht (2004), a difference in acquisition of archiphoneme /R/ in the several positions was not observed. But for Mezzomo (2005), appearance and acquisition of that archiphoneme in middle position happens after a stable end position.

The organization established for the acquisition of the coda structure by the Brazilian Portuguese reinforces the idea that the acquisition happens on a distinct way when the segment is in a simple structure (CV) and when it is in a more complex syllabic structure (CVC). Having as an example the case of the co-occurrences responsible for the non-lateral liquids, which will appear on the system according to Lazzarotto-Volcão (2009) at the age of about 3:4 years while on the coda position this segment starts appearing at the age of about 2:10 years and its acquisition is expected to happen at the age of 3:10 years. Such difference demonstrates to what the acquisition is sensitive to the extra segmental properties, as studies that assume that a same phonetic segment will show distinct formats of production in different contexts highlight (Freitas, Almeida & Costa, 2009; Almeida, Costa & Freitas, 2010; Matzenauer, 2013).

As such, we suggest four stages for coda acquisition. In the first, there are sonorant segments in end position. In the second, the obstruent and the middle position appear. In the third, all coda have been acquired, except archiphoneme /R/ structures, to be acquired in the last stage. This acquisition period goes until approximately 3:10 years. In that period, we highlight the relevance of structure position in the word, which affects acquisition periods. The hypothesis for a Standard Model of Structures and Contrast Acquisition for Coda can be seen in Figure 3.

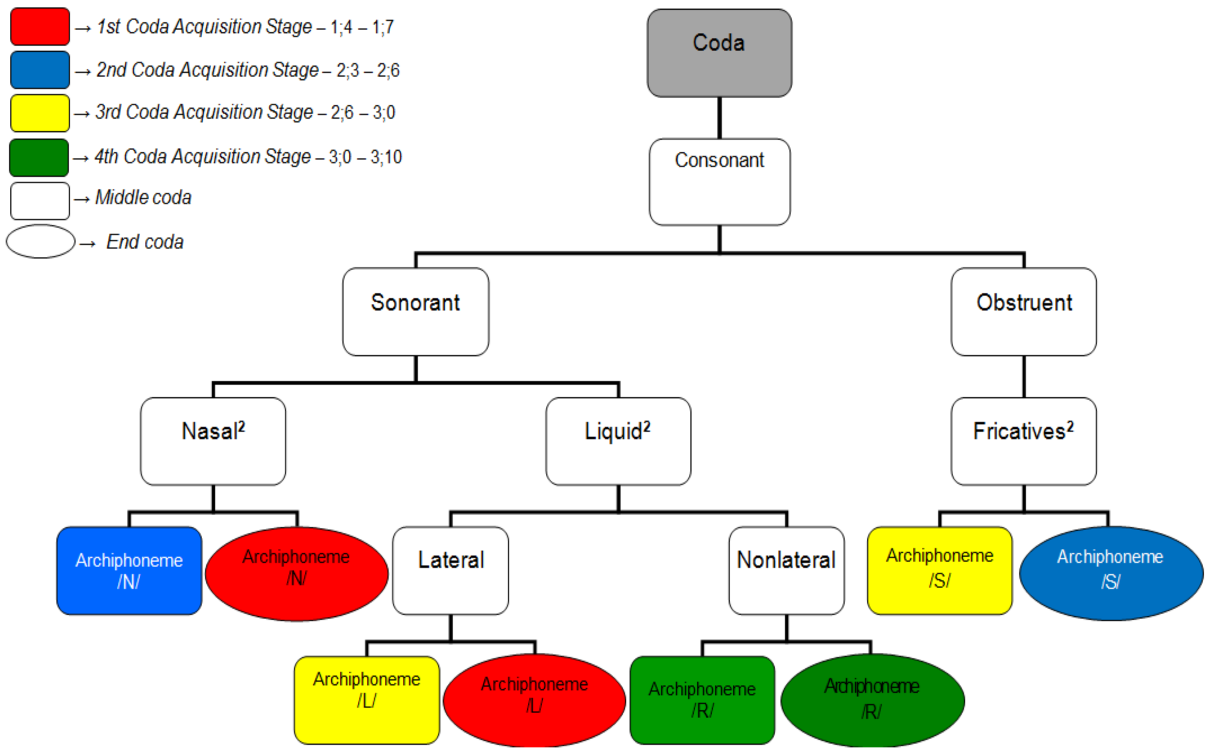


Figure 3. Standard Model of Structures and Contrast Acquisition for Coda

We point out that the configuration of onset presentation on the chart is to show contrasts hierarchy but that is not the intended idea for coda syllable structure. By using the chart to show coda acquisition, our aim is to present the difference between acquisition time for several syllable positions - middle and end - and not contrast between them.

2.2.3. Proposal for a Standard Model of Structures and Contrast Acquisition –Complex Onset

As referred on the introduction of this article, the onset is not a mandatory constituent of the syllable, but when it is present it can be filled by one or two consonants. When the onset is present and is filled by two consonants it characterizes a complex onset, and on the Brazilian Portuguese it will have on the first consonant an obstruent and the second consonant will be a liquid. However there are some restrictions regarding to the segments that can occupy the complex onset, being the allowed obstruent: /p, b, t, d, k, g, f, v/ and the liquids /l, ɾ/.

On the Brazilian Portuguese not all the combinations between these two elements are possible and some of the complex onset groups present restrictions regarding to the position on the word. For instance, the group formed by /vɾ/ occurs only in middle position (i.e. livro); the group /vl/ is found in some names, borrowed from other languages (i.e. Vladimir); and the words that contain a group /tl/ are few and when present the onset is in middle position preceded by the vowel 'a' (i.e. atlas). Besides, a great difference on the amount of existing words with the structure formed by /l/ and by /ɾ/ is observed, being the complex onsets formed with less frequent lateral liquid, conferring a property marked to the words that contain this structure.

The hypothesis supported by the Robustness Scale for Co-occurrence of Consonant Features has shown different acquisition stages for complex onset constituent. According to this, groups comprised of fricative obstruents together with lateral liquids would be acquired before complex onsets comprised of plosives and non-lateral liquid. This behavior was not observed in Brazilian Portuguese data, as these show different complex onset groups are acquired in a similar way.

As for complex onset, a different acquisition order for complex onsets in initial and in middle positions was not observed. This shows that syllable structure position is not relevant in that context, unlike what can be seen for coda. With the data available at Lamprecht (2004), complex onset acquisition can be defined in two stages:

1st Complex Onset Acquisition Stage (red) – Only the first structure consonant is performed, regardless of context. Factors such as type of consonant in C^1, liquid filling C^2 or position in word do not affect complex onset acquisition process. In this first stage, the child performs a CCV structure as a CV simple onset.

2nd Complex Onset Acquisition Stage (blue) – Appearance of CCV syllable, without any intermediate stage between structure simplification (CV) and its correct performance (CCV). In this stage, no preferences are observed for different sound classes (plosives or fricatives) or for type of liquid (lateral or non-lateral). According to data shown in Lamprecht (2004), acquiring different complex onset groups happens simultaneously. Thus, complex onset acquisition might occur in two moments - C^1V simplification and correct C^1C^2V performance. Command of the second stage happens around 5:0 years.

Based on data from Lamprecht (2004), one can see two acquisition stages for complex onset. The initial stage, where only the first structure consonant is produced (C^1V), is shown in the diagram by two conjoined rectangles: the first represents C^1 and is filled by onset segments and the colour red; the second rectangle has no filling. The second stage, where complex onset is acquired, is shown in the diagram by filling both rectangles with blue, thus proving the structure has been acquired by the child's system. The Standard Model of Structures and Contrast Acquisition for Complex Onset can be seen in Figure 4.

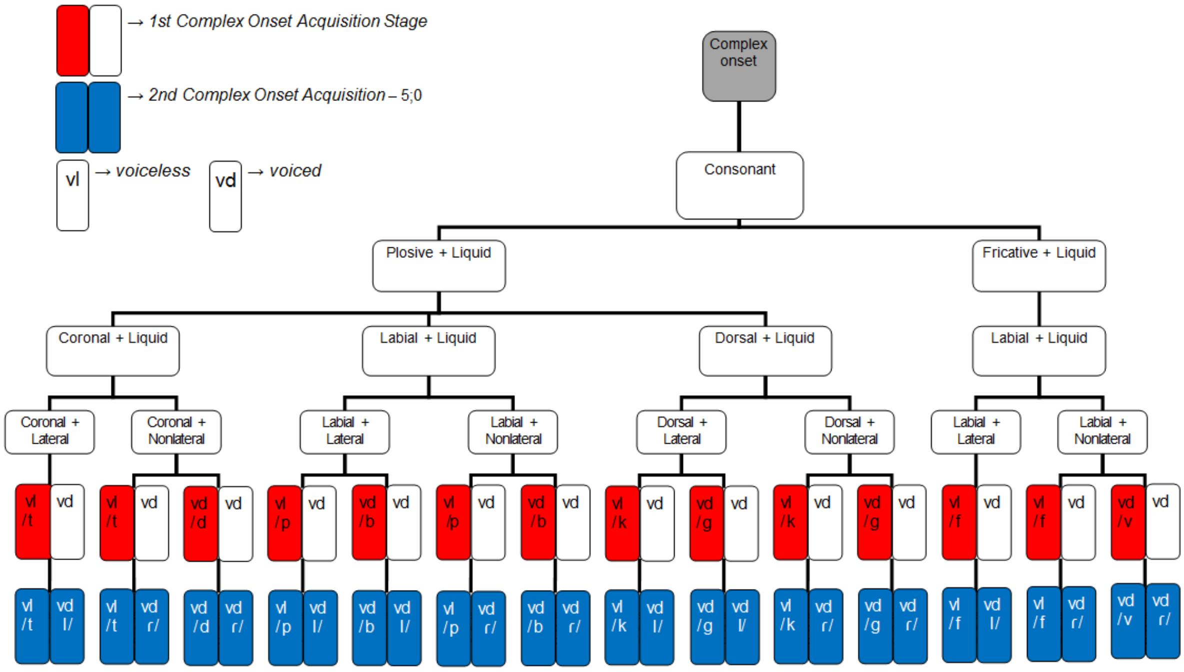


Figure 4. Standard Model of Structures and Contrast Acquisition for Complex Onset

One can see that the hypothesis based on Brazilian Portuguese empirical data is quite different from the first hypothesis supported by the Robustness Scale for Co-occurrence of Consonant Features. This difference refers to acquisition stages for different types of complex onset. Based on data from Lamprecht (2004), we conclude complex onset is acquired in the same way through all possible combinations for the structure.
The complex onset acquisition process on the Brazilian Portuguese can be better explained and justified through the Sonority Scale by Clements and Hume (1995). As referred previously, the syllable is formed by a sonority peak, the sonority is a natural property on the languages and is expressed from the degrees of prominence of this property, consequently from the segments that carry them. The sonority will determine the relation among the constituents that form the syllable, as well as the segments licensed to occupy each one of these constituents (onset, core, coda), and with this the sequences accepted on the language. The sonority scale characterizes as the distribution of the segments in a sequence due to the sonorant or sonority, as it can be observed on the Sonority Scale of Clements and Hume (1995) expressed on Table 3.

Table 3
Sonority Scale by Clements and Hume (1995)

Obstruentes	<	Nasal	<	Liquid	<	Vowel
0		1		2		3

In that rating scale, obstruents have the same value (0) and that would explain why complex onsets with plosives and fricatives are acquired simultaneously and treated in the same way by the child. Also, liquids, both lateral and non-laterals, have the same value (2), which explains why both /l/ onset and /ɾ/ onset are acquired at the same time. According to Clements (1990) the sonority profile preferred on the syllables presents a maximum sonority growth from the attack to the core and a minimum decrease from the core to the coda, which characterizes the complex onset standard on the Brazilian Portuguese.

Despite the complex onset formed by /l/ be in smaller amount and less frequent on the language and thus present a marked characteristic, what could justify a later acquisition (by the fact that the child tends to avoid more marked structures), it was not observed on the analyzed data of the Brazilian Portuguese.

The Robustness Scale for Co-occurrence of Consonant Features was created based on contrasts produced by different feature combinations observed at segmental level. By analysing the complex onset, we study the extra segmental, being such scale inappropriate to clarify the observed on the complex onset acquisition. It is known that the phonological acquisition depends not only on the features acquisition, but also on all the relation established among the different phonological units that compose the language.

For a better elucidation of the complex onset acquisition process, the Sonority Rating Scale by Clements & Hume (1995) is more suitable, as it presents equal sonority values for obstruents and for liquids, regardless of each one's class. In this way, it can be inferred that the children while acquiring the complex onset give great importance to the sonority cycle present on the syllable and they try to respect this property.

The greater ease (due to Strength) for acquiring plosive obstruents before fricative obstruents is observed when those segments happen in simple structures. When it comes to complex onset, the issue is no longer segmental but structural and sequential. As complex onset has a late acquisition (this structure is acquired around 5:0 years), the subjects segmental inventory is already complete and it is important to resort to other rating scales to explain the acquisition process.

The acquisition standard observed on the complex onset emphasizes the importance that the sonority has on the acquisition of this constituent. It is possible to observe that the child tries during the acquisition of the complex onset to keep the syllable sonority cycle (Clements, 1990), keeping the syllables always with a growing behavior from the onset to the core. It demonstrate that the phonological acquisition evolves the relation of the different phonological units of the language (for example, interaction between traces, syllabic constituent, tonicity and others), being the analysis of interaction among them very important.

3. Discussion and Final Comments

Spreading out the Standard Model of Contrast Acquisition, thus enabling an interaction between the syllable structure and the segmental inventory on phonological acquisition, allows a deeper analysis of children phonological systems. This makes it possible to determine syllable-segment interface in the presented speech data Brazilian Portuguese.

This extension has followed the same principles of its initial version, the Robustness Scale for Consonant Features (Clements, 2009), and of the Robustness Scale for Co-occurrence of Consonant Features (Lazzarotto-Volcão, 2009). From this initial analysis, it was observed that the phonological acquisition does not depend only on the acquisition and dominion of the features, but on the co-occurrence of features with other features already established on the system and thus to enable the acquisition of the language segments, as defended by Lazzarotto-Volcão (2009).

The scale proposed by Lazzarotto-Volcão (2009), with small changes, enabled to explain the segments acquisition, not only on the initial onset but also on the coda position. In spite of the fact that the set is a proposal based on the features acquisition, and their co-occurrence, the order proposed by the authoress was observed during the segments acquisition process on the simple onset and coda structures. This order did not seem to be valid to explain the steps of the segments acquisition on the complex onset position. To explain the segments acquisition on this context, the Sonority Rating Scale by Clements and Hume (1995) seemed to be more adequate. Enabling to check that the phonological acquisition of this constituent is guided by the syllable, because the child will acquire the complex onset only when the structure is available on its phonological system.

Based on these rating scales and on empirical data from Brazilian Portuguese, the following Standard Model of Structures and Contrast Acquisition was created, suggesting:

- Simple onset acquisition happens in four stages: stronger contrasts are acquired before weaker ones (as suggested by Lazzarotto-Volcão, 2009), Being that the acquisition occurs through co-occurrence among the features;
- Coda constituent acquisition also happens in four stages: end positions are acquired before middle positions, except when it comes to archiphoneme /R/, whose acquisition is co-occurrent.
- Complex onset acquisition happens in two stages: the first consonant of complex onset is acquired first and then happens the command of the structure, regardless of segment class.

This trend for segments in end coda position in Brazilian Portuguese to be acquired first and then to become stable in middle position agrees with other studies in this field showing the same conclusion (Lamprecht 1993; Miranda, 2012; Mezzomo, Dias & Vargas 2014; Mezzomo, Lopes, Abelin & Oppitz, 2015). However, this behavior is not observed on the European Portuguese speakers, recent studies showed difference on the acquisition process of liquids on the coda position (Lousada, Mendes, Valente & Hall, 2012; Amorim, 2014). According to the studies, the /l/ in coda is acquired firstly on the middle coda position and after on final coda, but the coda with

/ɾ/ initially is acquired in final position and after the stabilization on the middle position happens.

Study with children who are Brazilian Portuguese speakers with a typical phonological development comparing the acquisition process of the archiphonemes /R/ and /S/ on simple onset and coda position, verified that: for /S/ there was no difference between the acquisition on the different positions; the contrary is observed with /R/ where on onset position the children got more hits (Galea & Wertzner, 2010).

Dividing complex onset acquisition into two stages agrees with other texts from this field (Lamprecht, 1993; Ribas, 2003; Giacchini, Mota & Mezzomo, 2015; Lopes, Dias, & Mezzomo, 2015) showing absence of intermediate stages between complex onset simplification and its correct performance. This hypothesis does not show a tendency for acquiring either onset types instead of others, as verified by Ribas (2006). According to this hypothesis, every onset is acquired simultaneously by the subjects during typical phonological acquisition, strengthening the idea that complex onset stabilisation will happen when the child acquires the structure and not its isolated segments. The analysis strengthens the dependency relationship between the segmental and syllabic acquisition, as evidenced in other studies (Freitas, 1998, Freitas, Miguel & Faria, 2001; Ribas, 2006; Lamprecht, 2004; Almeida, Costa & Freitas, 2010; Matzenauer, 2013; Lopes, Dias, & Mezzomo, 2015).

The analysis performed from the Standard Model of Contrast Acquisition and Structure enables to verify which the problematic contrasts for the children in simple onset are, and also to check the interaction on these segments acquisition on the other syllabic structures. From the analysis of this interaction, it can be verified if the child´s difficulty is on the stabilization of some co-occurrence responsible for a determined segment, or if the problem is o the syllabic structure not available on the child´s system yet. This way, the Standard Model of Contrast Acquisition and Structure enables to evaluate the specific difficulties that the children can have in dealing with more complex structures. So, the description of the syllable-segment acquisition, reinforces the importance of the original model amplification (Lazzarotto-Volcão, 2009), since it evolutes the acquisition process in each one of the syllabic structures, checking if the difficulty is in the establishment of the contrasts and/or in the syllabic structures acquisition.

The description gathered by the Standard Model of Contrast Acquisition and Structure enables the Speech-Language Therapist to see if there are difficulties only at segmental level, only at prosodical level, or in both. From this it is possible to evaluate the gradual construction of systems, from the acquisition of simpler structures, to the domain of more complex structures. This kind of detail allows the therapist to perform a deeper assessment, to plan a more targeted treatment, and to offer a more efficient Speech-Language therapy. One must highlight the need for research using this model in a clinical environment, for analysing and selecting targets, to assess its applicability during the treatment process.

References

Almeida, L. (2011) Acquisition de la structure syllabique en contexte de bilinguisme simultané portugais-français. Doctoral thesis, Lisbon University, Lisbon.

Almeida, L., Costa, T. & Freitas, M. J. (2010) Estas portas e janelas: o caso das sibilantes na aquisição do português europeu. In: Brito, A. M., Silva, F., Veloso, J. & Fiéis, A. (eds) Textos Selecionados XXV Encontro Nacional da APL, 153-167.

Amorim, C. (2014). A aquisição das consoantes líquidas em português europeu: contributos para a caracterização da faixa etária 4;0 - 4;11 anos. *Revista de Estudos Linguísticos da Universidade do Porto, 9,* 59-82.

Boersma, P. (2007). Cue constraints and their interactions in phonological perception and production. *Rutgers Optimality Archive* 944.

Boersma, P. (2008) Emergent ranking of faithfulness explains markedness and licensing by cue. *Rutgers Optimality Archive* 954.

Camara, Jr., J. M. (1973) Princípios de Linguística Geral. Rio de Janeiro: Acadêmica.

Clements, G.N. The Role of the sonority cycle in core syllabification. In: Kingston, J.

& Beckman, M.E. (eds) Papers in laboratory phonology I. Between the grammar and physics of speech. Cambridge: Cambridge University Press, 1990. p. 283-333.

Clements, G. N. & Hume, E. V. (1995). The internal organization of speech sounds. In: Goldsmith, J. A. (org.) *The handbook of phonological theory.* Cambridge: Blackwell, 245-306.

Clements, G. N. (2009) The Role of Feature in Phonological Inventories. In: Raimy, E.; Cairns, C. E. *Contemporary Views on Architecture and Representations in Phonology.* Cambridge: MIT Press, 19-69.

Fikkert, P. (2007). Acquiring phonology. In: Lacy, P. (Org.). *Handbook of phonological theory.* Cambridge: Cambridge University Press, 537-554.

Freitas, M. J. (1998). Os segmentos que estão nas sílabas que as crianças produzem: localidade silábica e hierárquica de aquisição. In: *Actas XIII Encontro Nacional da Associação Portuguesa de Linguística,* Lisboa. 313.

Freitas, M. J., Miguel, M. & Faria, I. (2001). Interaction between Prosody and Morphosyntax: plurals within codas in the acquisition of European Portuguese. In: Hoehle, B. & Weissenborn, J.(eds.) *Approaches to Bootstrapping. Phonological, Lexical, Syntactic and Neurological Aspects of Early language Acquisition.* Amsterdam: John Benjamins Publishers 2, 45-58.

Galea, D. E. S. & Wertzner H. F. (2010). Comparison between syllabic onset and coda during phonological acquisition. *Revista da Sociedade Brasileira de Fonoaudiologia, 15,* 103-107. doi.org/10.1590/S1516-80342010000100017.

Giacchini, V., Mota, H. B. & Mezzomo, C. L. (2015). The details in the therapeutic process to the consonant clusters acquisition in the speech of children with phonological disorder. *Rev. CEFAC [online], 17,* 17-26. doi.org/10.1590/s1516-18462012005000049.

Grunwell, P. (1988). Phonological assessment, evaluation and explanation of speech disorders in children. *Clinical Linguistics and Phonetics, 2*, 221-252.

Grunwell, P. (1990) Os desvios fonológicos numa perspectiva linguística. In: Yavas, M. S. Desvios fonológicos em crianças. Teoria, pesquisa e tratamento. Porto Alegre: Mercado Aberto.

Henrich, V. & Ribas, L. P. (2014). Aquisição fonológica atípica do português brasileiro: relações implicacionais e de marcação na emergência das consoantes róticas. *Letrônica, 7*, 678-694.

Lamprecht, R. R. (1993). A aquisição da fonologia do português na faixa etária dos 2:9 a 5:5. *Letras de Hoje. 28*, 99-106.

Lamprecht, R. R. (org) (2004). *Aquisição fonológica do português: perfil de desenvolvimento e subsídios para a terapia*. Porto Alegre: Artmed.

Lazzarotto, C. (2007). A variação na aquisição fonológica sob a ótica da Teoria da Otimidade. *ReVEL, Special edition 1*, 1-12.

Lazzarotto-Volcão, C. (2009) Modelo padrão de aquisição de contrastes: Uma proposta de avaliação e classificação dos Desvios Fonológicos. Doctoral thesis, Catholic University of Pelotas, Pelotas.

Lazzarotto-Volcão, C. (2010). Uma proposta de Escala de Robustez para a aquisição fonológica do PB. *Letrônica*, 3, 62-80.

Lazzarotto-Volcão, C. (2012). O Modelo Padrão de Aquisição de Contrastes: Uma nova abordagem para o desvio fonológico. *Veredas OnLine – Especial*, 109-121.

Lopes, S. G., Dias, R. F. & Mezzomo, C. L. (2015) The production of the complex onset: acquisition guided by the syllable or segment?.*Rev. CEFAC[online], 17,* 78-87. doi.org/10.1590/s1516-18462012005000049

Lousada, M. Mendes, A. P., Valente, A. R. & Hall, A. (2012). Standardization of a Phonetic-Phonological Test for European-Portuguese Children. *Folia Phoniatrica y Logopaedica*, *64*, 151–156. doi: 10.1159/000264712.

Matzenauer, C. L. B. (2012). A teoria fonológica em estudos sobre a aquisição da linguagem. *Veredas Online, Special edition1,* 67-84.

Matzenauer, C. L. B. (2013). Unidades da fonologia na Aquisição da linguagem. *Revista ProlÍngua*, *8*, 17-40.

Mezzomo, C. L. (2005). O status das consoantes pós-vocálicas no PB: uma comparação com aquisição do holandês, sob enfoque da teoria de Princípios e Parâmetros. *Cad Pesqui Ling, 1,* 191-195.

Mezzomo, C. L., Dias, R. F. & Vargas, D. Z. (2014). Intervening factors to the correct production of (c)vc syllables in typical and atypical speech data. *D.E.L.T.A., 30*, 353-370. doi.org/10.1590/0102-445012167814494209.

Mezzomo, C. L., Giacchini, V, Dias, R. F., Luiz, S. W. & Lopes, S. G. (2011). The acquisition of simple and complex coda /S/ in children with phonological disorder. *Rev. CEFAC, 15*, 17-24. doi.org/ 10.1590/S1516-18462011005000128.

Mezzomo, C. L., Lopes, S. G., Abelin, C. M. & Oppitz, S. J. (2015). Acquisition of CV silable in the typical and impaired phonological development. *Rev. CEFAC [online], 17,* 60-71. doi.org/10.1590/1982-0216201517s14713.

Mezzomo, C. L., Quintas, V. G. Savoldi, A. & Bruno, L. B. (2010). Coda acquisition: a comparative study between transversal and longitudinal data. *Rev Soc Bras Fonoaudiol 15*, 401-407. doi.org/10.1590/S1516-80342010000300015.

Miranda, A. R. M. (2012). Reflexões sobre a fonologia e a aquisição da linguagem oral e escrita. *Veredas online, Special edition 1*, 122-140.

Montenegro, A. C. A (2012). Aquisição do onset complexo C(r) no português: uma abordagem multidimensional. Doctoral thesis, Federal University of Pernambuco, Recife.

Mota, H.B. (2001) Terapia fonoaudiológica para os desvios fonológicos. Rio de Janeiro: Revinter.

Othero, G. A. (2005). Processos fonológicos na aquisição da linguagem pela criança. *ReVEL, 3,* 1-13.

Ribas, L.P. (2003) Onset complexo: características da aquisição. *Letras de Hoje, 38*, 23-31.

Ribas L.P. (2006) Onset complexo nos desvios fonológicos: descrições, implicações para a teoria, contribuições para a terapia. [tese] Doctoral thesis, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre.

Rose, Y. (2000) Headedness and Prosodic Licencing in the L1 Acquisition of Phonology. Doctoral thesis, McGill University, Montreal.

Selkirk, E. (1982) The syllable. In: Hulst, H.V. D. & Smith, N. (org) The structural of phonological representations. Dordrecht: Foris.

Silva, M. K., Ferrante, C., Borsel, J. V. & Pereira, M. M. B. (2012). Aquisição fonológica do Português Brasileiro em crianças do Rio de Janeiro. *J. Soc. Bras. Fonoaudiol. 24*, 248-254. doi: 10.1590/S2179-64912012000300010.

Toreti, G. & Ribas, L. P. (2010). Aquisição fonológica: descrição longitudinal dos dados de fala de uma criança com desenvolvimento típico. *Letrônica,* 3 , 42-61.

Vitor, R. M. & Cardoso-Martins, C. (2007). Desenvolvimento fonológico de crianças pré-escolares da Região Noroeste de Belo Horizonte. *Psicol. rev., 13*, 383-398.

Wertzner, H., Pagan, L., Galea, D. & Papp, A. (2007). Características fonológicas de crianças com transtorno fonológico com e sem histórico de otite média. *Revista da Sociedade Brasileira de Fonoaudiologia*, *12,* 41-7.

Wetzels, L. (1997). The lexical representation of nasality in Brazilian Portuguese. *Probus 9*, 203-232. doi.org/10.1515/prbs.1997.9.2.203.

Wetzels, L. (2000) Consoantes palatais como geminadas fonológicas no Português Brasileiro. *Rev. Est. Ling., 9*, 5-15. doi.org/ 10.17851/2237-2083.9.2.5-15.

Yamaguchi, N. (2012) Parcours d'acquisition des sons du langage chez deux enfants francofones. Doctoral thesis, Université Sorbonne Nouvelle, Paris.

Yavas, M., Hernandorena, C. L. & Lamprecht, R. R. (2001). *Avaliação Fonológica da Criança: Reeducação e terapia.* Porto Alegre: Artmed.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 98-112, 2016, September
ISSN: 2148-1997

First language acquisition: A case study of a three-year old Lebanese child

Received : 11.01.2016
Accepted : 10.09.2016
Published : 30.09.2016

Joel C. Meniado[1]
University of the Philippines

Abstract

This case study focuses on the process of first language acquisition of a 3-year old Lebanese child. It also analyzes the factors and other mechanisms that influence L1 acquisition. For the duration of almost four months, the researcher observed and recorded the subject's produced sounds, words, and sentences. He also observed how the learner interacts with various linguistic inputs to see how he internally processes them. Finally, he also observed him as he interacts with people to determine the levels of the various constructs of his communicative competence. Based on the findings of the study, the subject acquired his first language (Arabic) largely biologically (nature). The observed behavior evidently confirms that he has a well-functioning Language Acquisition Device (LAD) that mainly enables him to acquire the language he is exposed to. Subordinately at his age, he is also helped by the environment to activate his innate capacity to acquire the language. Observations and interviews with his parents confirm that he also acquired through imitation and learned through correction and reinforcement, analogy, and structured input. Language universals are also evident in his communicative outputs while at the same time exhibiting some basic styles, registers, and non-verbal communication. The learner also manifested some linguistic and communication difficulties which are strongly influenced by his environment. Clearly, nature and environment play significant roles in a child's first language acquisition; thus, they should be creatively capitalized by both parents and teachers in various phases of instruction to ensure prolific and meaningful language learning and development.

Keywords first language acquisition, case study, factors in language acquisition, stages of language acquisition, acquiring Arabic language

1. Introduction

Language of any kind is exceedingly intricate, yet it is utterly amazing how children at a very young age (0-5 years old) are able to acquire or master it in their very own. Children during their Linguistic Period (1-5 years old) subconsciously perform such complex tasks as conjoining sentences, asking questions, using appropriate pronouns, negating sentences, forming relative clauses, and inflecting verbs and nouns (Fromkin, Rodman, and Hyams, 2010). Though sometimes with some lapses or imperfections, they can orally respond to questions and participate in conversations and argumentations. Some smart kids can also tell whether an adult is lying or not. This scenario makes anyone who is particularly interested in language learning wonder

[1]Bio: Joel C. Meniado has doctorate degree in education and masters degree in language and literacy education. He has been an ESL/EFL lecturer in Southeast Asia and Middle East for 16 years now. His research interests include language acquisition, educational technology, assessment, and reading education. Correspondence should be sent to joel.meniado@upou.edu.ph.

how young children without going to school or undergoing any formal instruction learn such complex structures and functions of a language.

There are many linguists attempting to explain such humanly overwhelming phenomenon. Lenneberg, Chomsky, and McNeill (in Brown, 2007; Orillos, 1998; Erneling, 1993; Green and Peil, 2016; Islam, 2013; Heidar, 2012), for example, share common belief that language is a "specie-specific" behavior and that it is innate in every individual. Bloomfield, Skinner, Fries, Hymes, Halliday, and Oller (in Brown, 2007; Orillos, 1998; Johnson, 2004; Ingram, 1989; Tahriri, 2013; Islam, 2013; Heidar, 2012; Ipek, 2009), on the other hand, believe that language is acquired behavior through conditioning and reinforcement and that it is best learned through interaction in various situations and applications. Based on these opposing views, language acquisition is influenced and controlled by two powerful factors – heredity (nature) and environment (nurture). The underlying question then is that which one is more influential over the other.

The question "How is language acquired/learned?" is already a long-ago question that has reaped numerous answers and criticisms through the years. Several theories sprung up with an attempt to uncover such a seemingly mysterious question but failed to completely clear up some doubts and offshoot questions. Though theories generally prescribe explanations to phenomena, they fail to attribute or recognize different personal circumstances and individual differences. Theories of Chomsky, Skinner, Hymes, etc. may be true to some but they may not be true to all individuals in all circumstances, hence, this study.

This study, though with a unique feat to answer some more specific questions on language acquisition in a more specific situation, had a main purpose of validating or testing the truthfulness of several language acquisition theories and principles using a case of a 3 year-old Lebanese Muslim child learning Arabic as his first language. It tried to find out if nature outweighs nurture (or vice versa) or if the two work complimentarily in this child in his quest for language acquisition and learning.

Primarily, this study tried to answer the following questions:

1. How does the subject of the study acquire his first language? Does he acquire the language by himself, by imitation, correction and reinforcement, analogy, or structured input?
2. What are the factors affecting his language acquisition?
3. How does the subject exhibit progression in his language acquisition?
4. What difficulties the subject encounters in language acquisition?

Answers to these questions are expected to contribute to the existing body of knowledge on language acquisition and learning in a specific context. Further, results are also expected to help teachers, educators, and curriculum developers design more comprehensive language teaching programs responsive to broader needs and natures of language learners. Lastly, this study can help parents develop more understanding and awareness of their children's needs and interests so they can lead them towards better and more productive language learning.

How is L1 acquired?

This section tries to deal with the question "How is first language (L1) learned or acquired?" Does an individual acquire a language biologically or through input and reinforcement of the environment? By reviewing previous theories and researches in the field, the concept and process of first language acquisition are explained.

Brown (2007), Orillos (1998), Johnson (2004), Tahriri (2013), and Heidar (2012) outlined several theories on the nature of a language which indirectly imply how a language (L1 & L2) is acquired. The Behaviorist / Empiricist Theory by Leonard Bloomfield, B. F. Skinner, and Charles Fries postulates that language is acquired behavior through conditioning and reinforcement. Based on this theory, children learn language from the conditioning and reinforcement of their environment which involve family members, peers, teachers, and the society as a whole. Further, it believes that language learning is the result of imitation, practice, feedback on success, and habit formation (Brown, 2000).

The Rationalist/Cognitivist/Mentalistic Theory of Noam Chomsky, on the other hand, believes that language is mainly cognitive, mental process, and rule-governed (universal grammar) in deep and surface structures (Brown, 2007; Green & Piel, 2016; Orillos, 1998; Islam, 2013; Tahriri, 2013; Heidar, 2012). Chomsky's theory is strengthened by Eric Lenneberg (in Brown, 2000; Orillos, 1998) with his Nativist Theory claiming that language is innate or inherent to a man that he can learn it in his own without environmental intervention. Further, this theory believes that man possesses a Language Acquisition Device (LAD) that enables him to acquire a language in a natural way (Brown, 2000; Orillos, 1998). Another claim of this theory is that language development follows biological and chronological program which means that various grammatical features are acquired according to natural order or program (Orillos, 1998).

The Sociolinguistic Theory of Dell Hymes and Michael A. K. Halliday focuses on the functionality and applicability of a language in various settings (Ma, 2009; Orillos, 1998). This is also otherwise known as the Functional Theory / Interactionist Model in other literatures claiming that language development goes along with the dynamic interaction of cognitive development and the environment (Brown, 2007; Orillos, 1998; Young, 1999; Blomaert (2014); Bargaric, 2007; Wardhaugh, 2006). This means that interaction or communication with adults within the child's cognitive capacities provide rich inputs that can contribute to language development. In this theory, there is a sort of fusion between behaviorist and cognitivist views on language acquisition.

The case study under investigation was anchored on the belief that first language acquisition is a complex phenomenon that it involves cognitive processes, overt behavior and habit formation. Though cognitive processes cannot be observed fully, they can be manifested through frames of thinking and overt behaviors; thus, the study observed the subject's language acquisition process through his attitudes and behaviors.

While there are some theories explaining how first language acquisition happens, there are also other behavioristic language acquisition mechanisms that further shed light on how language learning exactly occurs. What are exactly the roles of imitation, correction and reinforcement, analogy, and structured input in language learning?

Can a child learn through imitation? According to Fromkin, Rodman, and Hyams (2010), imitation plays a role to some extent. They questioned that if children mainly learn through imitation, why is it then that some children who are exposed to correct utterances, pronunciations, or sentences still come up with mispronounced words and ungrammatical sentences? Despite the fact that sometimes parents speak slowly to allow their kids imitate the correct word or sentences, still kids speak the wrong ones.

On the role of correction and reinforcement in language learning, behaviorists believe that children learn to produce correct (grammatical) sentences when they are praised or corrected. Children believe it is the right thing when adults show some approval or recognition (positive reinforcement) on correctly uttered sentences. Similarly, they will also learn when adults make some corrections (negative reinforcement) on their ungrammatical sentences. Corrections can be done in many positive ways like recast to avoid embarrassment on the part of the children. In the study of Whitehurst and Valdez-Menchaca (1988), they found out that kids increase vocabulary, comprehension, and productive skills performances when they are differentially reinforced. In a related study of Goldstein (1984), he found out that corrected practices enhance language learning.

Do Children learn language through analogy? According to Fromkin, Rodman, and Hyams (2010), analogy can help in some instances but not in most cases. Children often hear their parents utter phrases or sentences and they use these as their models in creating their own. They substitute some words in the phrases or sentences. But sometimes this creates serious problems as the meanings may also change. This happens especially when the children do not know the rules governing the proper arrangement of words in sentences. In the study of Gentner and Namy (2006), they revealed that there is a strong evidence that analogy helps in language acquisition.

In terms of learning through structured input, do children learn best when adults speak to them in a very special, simplified way called “motherese” or “baby talk”? Studies show that though infants prefer to listen to motherese over normal adult speech, using motherese does not significantly affect the child’s language development (Whyatt, 1994). In many cultures, many adults do not do baby talk but still babies are able to speak like children around the world do.

The concepts of imitation, reinforcement and correction, analogy, and structured input are all observable behaviors underlining the role of environment in language acquisition. The current case study involved observation of such behaviors to its subject; thus, they are deemed relevant and important.

1.1 Process and stages of language acquisition

Given the theories and mechanisms involved in language acquisition, the next question would be, “Do children of different backgrounds and cultures worldwide undergo the same pattern or stages of language acquisition?” According to innateness theory, every person is equipped with language acquisition device or brain language center with parts responsible for language functions (Brown, 2007; Orillos, 1998; O'Grady, 2008). These include Broca's area responsible for speech mechanisms, Wernicke's area responsible for receiving, selecting, and comprehending words, and the Angular gyrus responsible for reading and writing. Having these parts enables a person to acquire a language at certain stages and patterns (Orillos, 1998).

Piaget (in Orillos, 1998) also outlined stages of cognitive development of a child which also correspond to stages of language development. These include sensori-motor stage (from age 0-2), pre-operational stage (from age 2-7), and operational stage (from age 7-16). According to Orillos (1998), a person is capable of abstraction at the operational stage, thus, the most conducive to language learning.

Fromkin, Rodman, and Hyams (2010) contend that first language acquisition among children is fast but not instantaneous. Normally, the pattern is that children / babies start with babbling, then acquiring first words, and then putting words together in sentences until they reach the virtual adult competence which takes around three to five years. They also noted that observations of children acquiring different languages reveal that the stages are similar, possibly universal.

In the study of Suwandi (2010), he mentioned that the most amazing stage of language acquisition among children is from the age of 0 to 3 years or more. He also noted that one of the remarkable things about first language acquisition is the high degree of similarity seen in the early language of children, which includes cooing, gurgling, and crying. Infants distinguish differences of subtle sounds like "ma' and "pa." By the end of first year, they understand familiar / repeated words and are able to produce a word or two words. By the age of two, most children can utter at least 50 different words and start forming simple sentences – telegraphic sentences for that matter, omitting the function words (Lightbown in Suwandi, 2010).

The Linguistic Period (1-5 years old) of a child presents remarkable cognitive and language development. At the end of 4th and 5th year, most children will have acquired basic grammatical structures, adult-like articulation, morphological construction, and express a range of pragmatic intentions (syntax and semantics) and understand adult's utterances. Through these ages, a child gradually learns and acquires the four linguistic competences: phonology, morphology, syntax, and discourse (Suwandi, 2010).

In the case study conducted by Salim and Mehawesh (2014), they traced the language development of a Jordanian Arabic-speaking child. Using observation and phonetic transcription methods, they found out that the child acquired Arabic language at the age of five. They also noted the following linguistic development stages observed in their subject: 1) Early Vocalizations; 2) The Holophrastic Stage; and 3) the First Sentences.

2. Methodology

2.1 Research Design

Since this study's main aim was to determine and describe how a child learns / acquires his first language, it deemed appropriate to utilize the case study research design. A case study is an "in depth study of a particular situation rather than a sweeping statistical survey. It is a method used to narrow down a very broad field of research into one easily researchable topic" (Shuttleworth, 2008).

2.2 Subject of the study

The subject of this case study is named Al Barraa. He is a 3-year old Muslim Lebanese born to a middle-class Arabic family in Saudi Arabia. He is the first child of a couple who are both educated up to Bachelor's degrees in Lebanon. He has a brother and a sister who are in their babyhood and infancy stages. His father is a manager of a company headquartered in Lebanon while his mother is just a plain housewife. A physically, mentally, and socially healthy boy, he loves playing toy cars, running inside the house, watching Youtube videos, and touching every strange thing he sees. Still on his way of acquiring his first language, he always tries to ask his mother or father the names of things, actions, or ideas he observes inside and outside the house. Though his father knows and speaks a little bit of French, Arabic is the main language spoken at home. He is mainly exposed to an environment where Arabic is spoken and written. Having his father a manager of a book publishing company, he is exposed to panoply of Arabic books which he tries sometimes to open just to see the pictures or illustrations.

As the son of the researcher's friend whom he privately teaches Business English course at the comfort of his own home, Al Baraa is always around whenever the researcher teaches his Dad, playing around with his toys, disturbing his Dad, and talking to him whenever he pleases. He speaks Arabic, basically, but it is surprising that he can also speak some English words and answer some basic English questions like "What's your name? How old are you? Where are you from? What's the name of your sister?" At his age, he can converse in basic Arabic, though according to his Dad, he speaks with grammatical errors. He is very lovely, friendly, and inquisitive and he also displays vivid characteristics of those learning a first language. Hence, he was chosen as the subject of this L1 case study.

2.3 Data Gathering Procedure

The researcher mainly gathered the data through observations, then supplemented with interviews and questionnaires. He observed how Al Baraa talked to his father and sister – his nature and behaviors in communicating using verbal and non-verbal modes. The researcher could only slightly understand the way Al Baraa communicates, his facial expressions, tone, voice, etc. As to the content and meaning of what the subject was saying, the researcher always asked Al Baraa's father to translate them for him. *(Al Baraa's father is good in speaking English)*. During observations, the researcher had always with him some outlines or checklist – the things he should focus on which were based on the theories and principles of language acquisition. If he had some concepts/ideas which were not observed during his visit, he would always interview the subject's father or ask the subject's mother through a questionnaire. The researcher is also learning basic Arabic, so he had opportunities to speak with / interview Al Baraa in Arabic. Al Baraa's responses were noted and included in the analysis.

2.4 Mode of Analysis

The data gathered through observation notes, interviews, and recordings were analyzed based on the theories and principles of language acquisition since these were the bases in constructing the observation checklists, interview guides, and questionnaires. The analysis of the data was anchored on the confirmation or rejection of language acquisition theories, principles, and previous researches. If the said theories or principles were observed or manifested in the interview, then they were considered theoretically sound and valid. Otherwise, they were considered unsound, thus, referred for further scrutiny. Observations outside the established theories or principles were taken as new knowledge and were recommended for further examination.

3. Findings and Discussion

3.1 How does the subject of the study acquire / learn his first language?

Based on the series of observations conducted, Al Baraa's acquisition of his first language is largely innate or biological in nature (Innateness Theory). His possession of a normal and functional human brain enables him to perceive / receive linguistic inputs and produce communicative outputs. His overt behaviors manifested in various linguistic tasks confirm Lenneberg's nativist theory, Chomsky's innateness hypothesis, and McNeill's Language Acquisition Device (LAD) (Orillos, 1998). For example, his possession of a human brain that is capable of processing inputs enables him to respond appropriately to a given communicative situation. It would then be impossible for him to understand or utter words without such central processing unit – the brain which contains specific parts with their corresponding linguistic functions. It was also observed that the subject learned some Arabic words and sentences (though sometimes faulty) without undergoing any formal instruction. The amount of his linguistic knowledge (vocabulary, grammar, phonology, syntax, etc.) considering his very young age can be surprising to someone. According to his father, he learned some words on his own. He also establishes his own linguistic codes and systems. This could reflect the naturalness of language acquisition and learning. Language acquisition happens as a person grows or matures, hence, Chomsky's innateness hypothesis (Orillos, 1998).

Looking through McNeill's LAD (Orillos, 1998), the researcher observed the presence of such linguistic characteristics that confirm Al Baraa's possession of the device. He can distinguish speech sounds from other sounds in the environment. Such is manifested when he responds to his father or mother when his name is called upon and does not when his sister's or brother's were called. Another manifestation was that when he can point out the objects (e.g. sister, brother, sofa, chair, TV, iPad, etc.) when the researcher asks him to point out. It means he can distinguish sounds and associate them into real visible objects. He has also the ability to organize linguistic events into various classes. For example he can distinguish nouns (names) and verbs (actions). When his father asks him to dance, he doesn't point out to the sofa or computer. Instead, he moves / shakes his body. It means that at his age, the word dance for him is an action, not a name. He can also discern if his father is angry based on the loudness of his voice and facial expressions. Also, he can classify events into past, present, or future. He is also aware that language is a system, a way of fulfilling his needs and desires and making himself clearer when he communicates to his significant others would help him get his needs and wants easier. Lastly, he asks questions when he does not know names of things, persons, or actions. It means that he engages himself in constant evaluation to improve his own linguistic system. He acknowledges his own weaknesses and shortcomings and uses them as his bases in improving his language.

Al Baraa is right-handed, so the dominant brain hemisphere is the left hemisphere – the part wherein the LAD is housed (Orillos, 1998). Though the researcher was not able to examine the LAD or the left hemisphere physically, he was able to determine if the parts are present and are functioning perfectly by observing the linguistic behaviors associated to those parts. Al Baraa speaks clearly except the sound of "r" which is normal at his age. He does not have any speech defect; thus, his Broca's area is working well. He can also hear and recognize sounds and symbols / letters and can respond fairly when he is being asked. He can understand words and utterances and can communicate meaningfully; thus, his Wernicke's area is also functioning well. Lastly, he can read and scribble Arabic letters and some words. It means to say that his Angular gyrus is functioning well.

In general, Al Baraa has the biological (innate) facilities that enable him to understand and utilize a language. Without these, it would be impossible for him to acquire, learn, and use language. No matter how strong the environment is if his body parts responsible for language learning are impaired, he cannot basically acquire any language. It goes without saying then that Chomsky's innateness theory (Orillos, 1998) has played a role in Al Baraa's language acquisition and development.

3.1.1 The Role of Imitation, Correction and Reinforcement, Analogy, and Structured Input

While recognizing the invaluable contribution of biological factors in language acquisition and learning, the researcher also underlines the significant role of environment in Al Baraa's language development. The literatures reviewed presented some concerns and issues on the roles of imitation, correction and reinforcement, analogy, and structured inputs on child's language learning. However, on Al Baraa's case, the researcher saw how these acquisition mechanisms worked out for his language development. Language learning is like learning any other skill through imitation. On the case of Al Baraa, he produces words or sentences based on what he hears and based on what he understands from adults' conversations he is exposed to. While sometimes he fails in imitating exactly due to late developments in speech mechanism, it is evident that he wanted to imitate the same but hindered by some mechanisms. In his mind he wanted to do the right thing but physiologically he could not. It showed that he learned something from imitating and he will even learn better and faster if his speech mechanisms are fully developed. During the series of observations, it was observed that Al Baraa often imitates his father in speaking. Interviews with the father corroborate the observation. He often teasingly imitates questions and answers to such questions. According to his

father, he loves to imitate but if he is the one imitated, he easily gets annoyed.
Though he loves imitating, not all that he imitates are correctly imitated. For example, his father says "Wen baytok?" (*Where is your house?*) but he says "Kambaytok?" (*How much is your house?*). His father would correct him by changing the wrong word with the correct one, and then by repeating the sentence / question slowly so that he can understand the words fully. After series of corrections and repetitions, Al Baraa can master the correct word/sentence. Clearly, he learns through correction and reinforcement.
In a series of observations conducted, it was also evident that Al Baraa does substitution of word/s in sentences to make or construct sentences of his own. He uses sentences from adults as his guide in constructing sentences. Though sometimes he replaces words inappropriately in a sense that the new word distorts the meaning of the original sentence, somehow he still learned especially the patterns in constructing sentences.
On the use of structured input (motherese or baby talk), his father admitted in the interview that he and his wife did some simplification of words and sentences when Al Barra was younger. They did some baby talking just to emphasize words and concepts. His father admitted further that Arabic is phonologically complex language and sometimes speaking complex sentences so fast might disorient or confuse a child. Thus, his father believes that simplifying the language somehow helped Al Baraa acquire / learn the language.

3.2 What are the factors affecting the subject's first language acquisition?

Latency Factors. Aside from Biological factors playing significantly in Al Baraa's language acquisition, there are also other factors considered. These include neurological, psychomotor, cognitive, and affective considerations (Orillos, 1998). At his age (3), it is already clear that language acquisition is controlled by his left hemisphere as manifested by his right-handedness. His inability to pronounce words with "r" sound is also explained by Krashen's (in Orillos, 1998) theory that lateralization is complete at the age of 5 and it is only after 5 that a child masters the authentic pronunciation. Since he is only 3, body parts and mechanisms responsible for the production of the "r" sound is not yet fully developed (psychomotor considerations).

Cognitive Factors. Based on Piaget's (in Orillos, 1998) stages of cognitive development, Al Baraa is still on the pre-operational stage (from age 2-7). This explains why his words are still limited. Normally, a person develops his ability for abstraction as he grows old, and Al Barra being at the age of 3 still has a lot of words to learn depending on his brain capacity and complexity. Using his age as a parameter in judging his cognitive ability, he is within the normal range.

Affective Factors. Al Baraa's personality seems to be an advantage in his first language acquisition. Based on the researcher's observations, no inhibitions, anxiety, or shyness were noted. He is very extrovert, friendly, and energetic, hence, having low affective filter. This kind of personality somehow has a positive impact on the pace of language acquisition. Krashen's (in Orillos, 1998) affective filter hypothesis states that when motivation and self-esteem are high and anxiety is low, the affective filter will also be low which means that the rate of language acquisition is faster.

Competence and Performance. Linguistic competence is one's internalized knowledge about the system of the language while performance is the observable and concrete manifestation or realization of competence (Orillos, 1998). Though it was hard to determine the level of competence of Al Baraa, the researcher looked into and measured his performances in the four (4) macro-linguistic skills. The researcher believes that a performance / linguistic output reflects competence, thus he tried to establish a point of leveling or categorizing to his subject of the study. For the sake of establishing the level of proficiency in various language skills, the researcher used his professional experience in leveling students. Based on the researcher's observation and evaluation, Al Baraa is at the Beginner's Level (A1). This is confirmed by his father who is a native speaker of Arabic. Looking closely on the four linguistic skills, vocabulary, and grammar, speaking seems to be his highest, followed by listening, vocabulary and grammar, then reading, and writing.
On Chomsky's (in Orillos, 1998) idea that children may already have linguistic competence even before they show linguistic performance, the researcher basing on his observation to his subject agrees to some extent. Al Baraa has the linguistic competence that is only appropriate at his age. Competence develops as a person grows and matures (Orillos, 1998). The younger the normal person is, the lower his competence in his native language. This is the usual pattern of language development.

3.3 How does the subject exhibit progression in his language acquisition?

Language Universals. Chomsky (in Orillos, 1998) believes that children from any nationality, background, or culture learn a language in the same manner and pattern. In the case of Al Baraa, at his age, he is already able to fairly manifest such universal linguistic categories as: nouns and noun classes, verb and verb classes, predication, negation, and question formation. On the other hand, he still has limited manifestations on word order, morphological marking, tone, agreement, and reduced reference.

Communicative Competence. Communicative competence is "the aspect of a person's competence that enables him to convey and interpret messages and to negotiate meanings interpersonally within specific contexts" (Orillos, 1998). It has four different components – the grammatical competence, discourse competence, sociolinguistic competence, and strategic competence.
Focusing on Al Baraa's case, evidently he does not yet have full knowledge of lexical items and rules of morphology, syntax, semantics, and phonology (*grammatical competence*). Though he can basically construct sentences on his own, his sentences are very basic. Sometimes the function words are omitted, thus, telegraphic sentences. When he uses analogy / substitution

in sentences of others as he makes his own, his choice of words is inappropriate or sometimes misplaced, thereby creating ambiguities in meaning.

On *discourse competence*, he can fairly comprehend 2-3 connected sentences in a conversation. One cannot expect him to respond meaningfully to utterances beyond three sentences. Thus, his discourse competence is still very low. On *sociolinguistic competence*, it was observed that he is on an average / intermediate level. He knows how to adjust his language depending on the person he is speaking with. He is more spontaneous when he speaks to his father and holds some reservations when he speaks to the researcher as a stranger. He can also discern if his father is angry or joking, tired, or interested to listen to him.

As a young child learning his first language, it is given that he has limited grammatical competence. Consequently, he finds difficulty in effectively expressing himself in a communicative situation. However, the researcher observed that he is able to use verbal and non-verbal communicative strategies (*strategic competence*) to compensate his difficulty in communicating. For example, when he wants something from his father but does not know its name, he can simply point to the object or illustrate it through action. In some instances, he also uses simpler, related words just to lead the listener to arrive at the message he wants to convey.

Styles and Registers. Orillos (1998) mentioned that native speakers as they mature into adulthood learn to adopt appropriate styles for widely different contexts. Al Baraa at his young age is able to vary his styles in communicating to different types of people. Of the five styles outlined by Joos (in Orillos, 1998), he can perform two – the casual and intimate styles. When he communicates with his parents and siblings he uses the intimate style. He speaks to them without reservations or inhibitions, revealing his true feelings and emotions. On the contrary, when he talks to people outside his family circle, he distances himself a bit. When he talks to the researcher, for example, he cannot whine, or speak with his tantrums. He answers straightforward though certain degree of trust and closeness can still be felt.

When it comes to his register, it was observed that he uses the informal Arabic often heard among youngsters on the streets and public places. Just like in English where there is called General English, Arabic language also has General Arabic. Due to his exposure to general conversations between and among his parents and relatives at home, he acquired and developed the General Arabic register which is different from formal Arabic used in schools and offices and the classical Arabic also known as Quranic Arabic which is used in the mosques and worship programs.

Nonverbal Communication. Nonverbal communication is indispensable in any communicative act. Even for infants or babies, they use crying to communicate to adults that they feel hungry or uncomfortable. For adults, they use smile, nod, or shaking of the head to show agreement or disapproval. In the case of Al Baraa, the researcher noted several types of non-verbal communication most especially when the subject talks to his father or to the researcher himself. For example, he does not look at the person he is talking to if he is disinterested or bored. He keeps on walking while talking if he is not interested with the topic. If he meets strangers, he distances himself. If he knows his father is angry, he does not talk or does not show up at all even if he is called upon. When he meets new visitors, he always carries with him his new toys or gadgets (e.g. cars, iPad, etc.). He has the attitude of dominance thinking that showing his new toys or latest gadgets to others would give him an elevated socio-economic status. When he already developed closeness with the researcher (after some time), he began touching his belongings like laptop, notes, books, food, etc. This means that the barrier was already broken. When he asks some help from the researcher, he approaches him closely and even holds him on his hands or sits on his lap. This means that he needs something badly or urgently.

3.4 What difficulties the subject encounters in language learning / acquisition?

Difficulties in Language Learning. Basically, Al Baraa's language learning difficulties include grammatical and discoursal aspects of the language. He still has limited vocabulary, phonological / mispronunciation problems, grammatical lapses particularly in arranging / placing appropriate words in a sentence, and combining sentences to produce meaningful discourse. Since based on observations nature has endowed him with brain capable of processing and discharging linguistic functions making his LAD quite working well, such difficulties are largely attributed to environmental forces and mechanisms. Due to his young age, his exposure to the language is still limited; thus, his communicative competence spread on various constructs is still low using the adult proficiency as the parameter. However, if he is compared to other children his age, he is not considered a problematic child when it comes to language acquisition and learning.

4. Conclusions

First language acquisition is shaped and influenced mainly by two factors: biological and environmental factors. As manifested in the case of the subject in this study, he basically acquired his first language with the presence and aid of his Language Acquisition Device (LAD). Without such brain facility, it would have been impossible for him to acquire his first language – the Arabic language. Several considerations also played in the natural acquisition of his language – neurological, psychomotor, cognitive, and affective. The maturation of his brain parts responsible in discharging linguistic functions, the development of his muscles responsible in the production of speech sounds, and the elimination of his emotional filters that hinder his language acquisition contributed to the natural evolution of his first language. Though his ability to acquire his first language seemed innate, its development and enhancement were further shaped by the environment. These environmental forces came in the form of imitation, correction and reinforcement, analogy, and structured input.

The period of the case study was such a limited time. However, the researcher noted some patterns of development in the subject's language acquisition and learning. He exhibited characteristics of Jean Piaget's (in Orillos, 1998) stages of cognitive and language development. Further, he also manifested that as he grows older, the more complex his language becomes. It shows that one's mastery or proficiency of a language goes with his maturity and extensive exposure with the language.

The subject of the study also manifested some challenges and language difficulties. These difficulties were not largely caused by biological factors; instead, these were created by the environmental factors. His inability to pronounce some words perfectly (phonology), arrange words in a complete sentence (syntax), combine sentences meaningfully (discourse), and understand word/sentential meanings in various levels (lexicon and semantics) was more of an environmental issue caused by his still limited exposure with the language. This reasoning would conclude that language acquisition and learning as well as development of communicative competence is affected by the environment.

Given that nature and environment interplay in the language acquisition process, it is important that language teachers recognize and consider individual needs and circumstances in planning their lessons. For example, they could administer multiple intelligences or learning styles inventory to identify the strengths and weaknesses of their learners in learning. Neurological, psychomotor, cognitive, and affective developments may cause delayed language development or poor linguistic performance / competence. Knowing how these considerations affect a child's language acquisition will help the teacher plan and execute his lessons well.

Since it was also noted that environment influences a child's first language acquisition, it is imperative that both parents (the child's first teachers) and school teachers provide a conducive and supportive environment where the child can easily acquire and fully learn a language. For example, schools can adopt and implement learner-centered pedagogies, facilities, pograms, and policies. Parents and teachers should also serve as models of grammatically correct language, so children could imitate the correct ones. Also, since correction and reinforcement are found effective in language learning, parents and teachers should learn the various ways of correcting and reinforcing such as use of oral re-casts, direct corrective feedbacks, indirect corrective feedbacks, metalinguistic corrective feedbacks, and reformulation. Use of analogy and structured inputs could be devastating if not properly administered / monitored; hence, it is also recommended that use of such should be taken with extra care.

The development of communicative competence is the pinnacle of language teaching and learning; thus, this must always be the overarching goal of every parent and teacher in teaching language among children. They should always subscribe to communicative language teaching where children are exposed to authentic communicative situations. Children should always be given ample opportunities to be immersed with the language and to use the language through different communicative activities in and outside the home or classroom.

As there is an identified pattern of language acquisition and development among children, teachers must respect this course of nature in lesson planning and teaching. For example, in formulating lesson objectives and

designing enrichment and assessment activities, existing linguistic and cognitive abilities of learners should be considered. Respecting learners abilities would allow them to make their lessons not too easy and too difficult. Further, it would make their language teaching fun and prolific.
There may be some difficulties encountered by children in language acquisition / learning but with the teacher's thorough understanding on the factors (biological and environmental) causing such, it would be easier for him/her to develop antidotes that would lead learners fully acquire and learn the language. Language acquisition can be complex and challenging but with teacher's thorough awareness and proper training, it can be fun and interesting.

References

Bagaric, V., & Djigunovic, J. M. (2007). Defining communicative competence. *Metodika*, *8*(1), pp. 94-103. Retrieved from http://hrcak.srce.hr/file/42651.html

Blommaert, J. (2014). From mobility to complexity in sociolinguistic theory and method. In Coupland, N. (Ed.) *Sociolinguistics: Theoretical Debates*. Cambridge: Cambridge University Press. Retrieved from https://www.tilburguniversity.edu/upload/5ff19e97-9abc-45d0-8773-d2d8b0a9b0f8_TPCS_103_Blommaert.pdf

Brown, H. D. (2000). *Principles of language learning & teaching*. 4th ed. New York: Pearson Education.

Brown, H. D. (2007). *Principles of language learning and teaching*. 5th ed. New York: Pearson Education.

Erneling, E. E. (1993). *Understanding language acquisition: The framework of learning*. Albany: State University of New York Press.

Fromkin, V., Rodman, Rodman, R., and Hyams, N. (2010). *Introduction to Linguistics*. Philippines: Cengage Learning Asia.

Gentner, D., & Namy, L. L. (2016). Analogical processes in language learning. *Current Directions in Psychological Science*, 15(6), pp. 297-301. Retrieved from http://groups.psych.northwestern.edu/gentner/papers/GentnerNamy06.pdf

Goldstein, H. (1984). Effects of modeling and corrected practice in generative language learning of preschool children. Journal of Speech and Hearing Disorders, 49, pp.389-398. doi :10.1044/jshd.4904.389

Green, M., & Piel, J. (2016). *Theories of human development: A comparative approach* (2nd ed.). New York: Routledge.

Heidar, D. M. (2012). First language acquisition: Psychological considerations and epistemology. *Theory and Practice in Language Studies*, 2(*2*), pp. 411-416. doi:10.4304/tpls.2.2.411-416

Ingram, D. (1989). *First Language Acquisition: Method, Description, and Explanation*. Cambridge: Cambridge University Press.

Ipek, H. (2009). Comparing and contrasting first and second language acquisition: Implications for language teachers. *English Language Teaching*, *2*(2), pp. 155-163. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.2886&rep=rep1&type=pdf

Islam, M. T. (2013). First language acquisition theories and transition to SLA. In Ali, S., et al. (Eds.). Paper presented at the Asian Conference on Language Learning 2013 Official Conference Proceedings, Osaka, Japan (499-510). Aichi, Japan: The International Academic Forum (IAFOR). Retrieved from http://iafor.org/archives/offprints/acll2013-offprints/ACLL2013_0289.pdf

Johnson, M. (2004). *A Philosophy of second language acquisition*. New Haven: Yale University Press.

Ma, T. (2009). On communicative language teaching – theoretical foundations and principles. Asian Social Science, 5(4), pp. 40-45. doi: http://dx.doi.org/10.5539/ass.v5n4p40

O'Grady, W. (2008). Innateness, universal grammar, and emergentism. *Lingua*, *118*(2008), pp.620-631. doi:10.1016/j.lingua.2007.03.005

Orillos, L. Q. (1998). *Language acquisition theories, principles, and research.* Quezon City: UPOU.

Salim, J. A., & Mehawesh, M. (2014). Stages in language acquisition: A case study. *English Language and Literature Studies, 4*(4), 16-24. doi:10.5539/ells.v4n4p16

Shuttleworth, M. (2008). *Case Study Research Design*. Retrieved from http://www.experiment-resources.com/case-study-researchdesign.html

Suwandi, S. (2010). A Thorough study on a child learning her first language: A case study on a three-year old child. *Eternal, 1* (2), 1-15. Retrieved from http://e-jurnal.ikippgrismg.ac.id/index.php/eternal/article/view/178/169

Tahriri, A. (2013). Revisiting first language acquisition through empirical and rational perspectives. *International Journal of Social Science & Education*, *3*(3), pp. 677-682. Retrieved from http://ijsse.com/sites/default/files/issues/2013/v3i3/Paper-14.pdf

Wardhaugh, R. (2006). *An introduction to sociolinguistics*. 5th ed. Massachusetts: Blackwell Publishing.

Whitehurst, G. J., & Valdez-Menchaca, M. C. (1988). What is the role of reinforcement in early language acquisition? Child Development, 59(2), pp. 430-440. Doi: 10.2307/1130322

Whyatt, B. (1994). Baby talk – The language addressed to language-acquiring children: a review of the problem. Studia Anglica Posnaneniensia, 29, pp. 126-135. Retrieved from http://www.ifa.amu.edu.pl/sap/files/29/10Whyatt.pdf

Young, R. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics*, 19, pp. 105-132. Retrieved from http://www.english.wisc.edu/rfyoung/young1999.pdf

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

Agreement groups coverage of English mother-child utterances for modelling linguistic generalisations

László Drienkó[1]
ELTE University

Received : 20.08.2015
Accepted : 08.08.2016
Published : 30.09.2016

Abstract

The present paper summarises findings from a series of experiments using the 'agreement groups' method, a distributional framework for analysing linguistic data. First, the method was applied to short mother-child utterances in order to directly investigate how group formation can affect language processing. Next, it was examined how longer utterances could be processed with the help of the groups of the previous analyses, i.e. how utterance fragments compatible with 'agreement groups' could 'cover' longer utterances. After recapitulating our previous results from English, Hungarian, and Spanish analyses, we report our findings on the 'coverage' of English utterances. Furthermore, we extend the coverage mechanism by employing discontinuous fragments and point out some theoretical implications, outlining a formal "continuum" model of linguistic generalisation. Convergence points with usage-based and constructionist approaches will also be discussed. Our method is "computationist" in that it emphasises the computational aspects of linguistic processing.

Keywords agreement groups, language acquisition, mother-child utterances, linguistic generalisation

1. Introduction

The 'agreement groups' framework for the distributional analysis of linguistic datasets was first proposed in Drienkó (2012a). The analysis is based on groups of minimally contrasting utterances. Such groups can also be seen as representing underlying agreement relations (as understood in the generalised sense of Drienkó 2012b). Agreement groups provide a means for processing novel utterances on the basis of utterances already encountered. Analysing 2-5 word long English mother-child utterances, Drienkó (2012a) found that at any stage of linguistic development some percentage of novel utterances were compatible with the agreement groups extracted from the body of utterances encountered that far. The results were slightly improved when a "category guessing" mechanism was added. The author also claimed that the formation of groups may support categorisation, and the actual emergence of grammatical agreement. Similar results were reported in Drienkó (2013a) for Hungarian and Spanish. Drienkó (2013b) extended agreement groups analysis for sequences larger than the utterances in the groups: it was investigated how agreement groups can account for fragments of longer English utterances, i.e. to what extent longer utterances can be "covered" by agreement groups. The findings of a similar investigation in the context of Hungarian mother-child discourse were reported in Drienkó (2013c). The present paper summarises previous findings from agreement groups experiments, with more specific focus on English coverage results and extends the coverage mechanism by employing

[1] Bio: László Drienkó studies Language and Cognition and Cognitive Linguistics at ELTE University- Hungary. drienko@t-nline.hu

discontinuous fragments. Also, some theoretical implications will be emphasised in the context of linguistic generalisations that the agreement group framework supports. Finally, we establish links to usage-based and constructionist research.

2. Methodology

2.1. Agreement groups as distributional analysis[2]

2.1.1. Agreement groups

Generally speaking, an agreement group is a group of symbol sequences where each sequence differs in exactly one symbolic unit from a given 'base' sequence, and where all sequences are supposed to be legal, i.e. compatible with some sequence-creating mechanism. In the context of natural language syntax this means that an agreement group consists of a grammatical 'base' utterance plus its minimal pairs, i.e. grammatical utterances differing in exactly one word from it. Naturally, the conditions for forming such a group of utterances of any language are determined by the distributional properties of the language in question.

Intuitively, our "one-word difference" criterion can be brought into connection with the notion of grammatical agreement. That is why we refer to a base utterance together with its minimal pairs as an agreement group. Our logic in relating "one-word difference" to grammatical agreement can be outlined as follows.

If we replace one word in a grammatical utterance with a word of the same "lexical category", and if the new utterance is grammatical, then we may assume that agreement relations must have been preserved since the new word fulfils the agreement requirements of the original sentence. Furthermore, for unambiguous cases we expect the preservation of the original agreement values. For instance, in example (1) below we replace the noun 'Adam' with the nouns 'Eve' and 'people' – cf. the boldface elements – and get a grammatical and an ungrammatical sentence, respectively. In the 'Eve' case the resultant sentence is grammatical and the agreement feature values, 3rd person and singular, are preserved. In the 'people' case the resultant sentence is ungrammatical so we do not expect agreement relations to have been preserved. Actually, it is the mismatch of number values – singular versus plural – that causes ungrammaticality[3]

(1)

Adam hates football → **Eve** hates football (3rd person singular agreement between subject and verb preserved)

→ ***People** hates football (agreement relation spoilt)

[2] This section is an adapted version of the introduction in Drienkó (2014).

[3] Of course, if we consider 'people' to belong to a different lexical category – common noun as opposed to the proper nouns 'Adam' and 'Eve'– we do not expect agreement feature matching in the first place since the "same lexical category" replacement requirement is not fulfilled. We would not like to go into much detail about the possibility of a precise definition for lexical categories. An intuitive notion of lexical, or grammatical, category will suffice in setting the context for our findings.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

When more words are involved in the change, the preservation of agreement values cannot be guaranteed. The replacement of 'Adam' and 'hates' in 'Adam hates football', for example, does not affect agreement features for 'Eve likes football', however in 'People hate football' the values are different[4]. Cf. (2).

(2)
Adam hates football → **Eve likes** football (agreement values preserved: 3rd person singular)

→ **People hate** football (agreement relation between subject and verb is preserved, but the 3rd-singular is changed to 3rd-plural)

In the case of words with ambiguous lexical categories agreement relations will be preserved when an appropriate lexical item is substituted, i.e. a lexical item whose agreement features are compatible with the requirements of the context. In the Hungarian examples of (3) below, the lexically ambiguous word *vár* 'castle/waits' can be replaced with the verb *létezik* 'exists' without changing the subject-verb person and number agreement relation. Similarly, when *vár* is replaced with an appropriate noun, e.g. *szentély* 'temple', the obligatory number agreement between the subject and the predicate noun is maintained with the same values, i.e. singular for both.

(3)

Az iskola **vár** (the school waits/is waiting 'the school is waiting(for you)')	→	Az iskola **létezik** (the school exists/is existing 'the school exists')
Az iskola **vár** (the school castle 'the school is a castle')	→	Az iskola **szentély** (the school temple 'the school is a temple')

In the case of words with ambiguous agreement values the actual values might change, but the agreement relation is still preserved. In (4), for instance, the noun 'sheep' can be considered both plural and singular. In a singular context it can be replaced with a singular noun, e.g. 'dog', in a plural context with a plural noun, e.g. 'dogs' – agreement values trivially will not change. When 'sheep' contributes to the context, however, a word change can alter the actual agreement value, as does the substitution of 'are' for 'is'. Nevertheless, the agreement relation between 'sheep' and 'are' is still correct.

(4)
The **sheep** is there→ The **dog** is there (3rd person singular agreement between subject and 'be' preserved)

---

[4] For simplicity, we regard verbs without an '-s' suffix as 3rd person plural.

The **sheep** are there→ The **dogs** are there (3rd person plural agreement between subject and 'be' preserved)

The sheep **is** there→ The sheep **are** there (agreement relation between subject and verb is preserved, but the 3rd-singular value is changed to 3rd-plural)

The above examples were meant to suggest that the agreement properties of "minimal pairs" of grammatical utterances, obtained by changing one word in an utterance, are fairly similar.
Next we observe that for a given utterance several such minimal pairs can be formed. Consider the sentences in (5).

(5) Adam hates football
Adam hates **basketball**
Eve hates football
Adam **dislikes** football
Charles hates football

Here each sentence is obtained by altering exactly one word in the underlined 'base' sentence 'Adam hates football'. In each sentence there is an agreement relation between the subject and the verb, invariably with the values 3rd person, singular. Since such a group of sentences represents similar agreement properties[5], we may term it an 'agreement group'.

2.1.2. Tabular representation of agreement groups – novel sequences

Now let us arrange the words of the sentences of our agreement group in a table like below, where each word occurs only once. Cf. Table 1.

Table 1

Agreement group in tabular form

Adam	hates	football
Eve	dislikes	basketball
Charles		

Evidently, each sentence in the group can be composed by properly concatenating words from different columns of the table. We expect the words in the same column to have similar agreement properties since they belong to the same lexical category and yield grammatical sentences when substituted into the original utterance 'Adam hates football'. Accordingly, we can say that words within a column belong to the same "agreement category", and we expect any new combination of words, as dictated by their agreement categories (columns in the table), to be grammatical.

[5] In the general case, by 'agreement' we refer to the generalised notion of 'agreement' proposed in Drienkó (2004a, b; 2012b). In that sense, feature correlation of any nature – phonological, morphological, even configurational – can be regarded as an agreement relation.

Indeed, we observe that the seven novel sentences in (6) are also compatible with the tabular representation of the agreement group, and that agreement features do not change.[6]

(6)

Eve hates basketball	Eve dislikes football	Eve dislikes basketball
Charles hates basketball	Charles dislikes football	Charles dislikes basketball
Adam dislikes basketball		

Table 1 was created on the basis of five sentences, which means that an originally 5-member group represents 5+7=12 sentences altogether. As sentences grow longer – more columns in the table –, and vocabulary becomes larger – more words in the columns (i.e. more rows) –, the "generative power" of agreement groups becomes more impressive. For instance, 401 sentences may suffice for the creation of a table with 4 columns and 101 rows, representing $101^4 \approx 10^8$ sentences altogether. Thus, besides carrying immediate information about the grammatical build-up of a language, agreement groups can provide a basis for processing novel sequences. This latter property renders agreement groups a useful tool for modelling processes characteristically involving change over time, notably learning processes, where a training phase typically precedes a given state of competence.

2.1.3. Agreement-groups analysis

An agreement-groups analysis of a given corpus of linguistic sequences may proceed along the following basic lines:

1- Extract all possible agreement groups from a body of training sequences.
2- Assign to every word in the training corpus at least one agreement category, where agreement categories are determined by the columns of (hypothetical) tabular representations of agreement groups.
3- Calculate the proportion of novel sequences in the test corpus that are 'compatible' with the agreement groups gained from the training corpus.

Extracting all possible agreement groups may naturally result in redundancy since there may be groups with identical members. Such groups can be

[6] Note the importance of our "one-word-difference" requirement for agreement groups. If, for instance, 'Adam hates football' and 'People hate football' belonged to the same group, the incorrect 'Adam hate football' and 'People hates football' would also be licensed by the corresponding table:

Adam	hates	football
people	hate	

eliminated during computation, however we believe that redundancy should be a component of a realistic model. Corpus a) in (7), for an example, yields agreement groups b) - f), of which e) and f) contain the same sentences, and b) consists of the sentences of c) and d).

(7)

a)	b)	c)	d)
Adam hates football	Adam hates football	Adam hates basketball	Eve hates football
Adam hates basketball	Adam hates basketball	Adam hates football	Adam hates football
Eve hates football	Eve hates football		
Boys hate football		e)	f)
Girls hate football		Boys hate football	Girls hate football
		Girls hate football	Boys hate football

Note that the mere inspection of agreement groups may reveal some regularities in the data. The first word in every sentence of (7) represents some human being(s), the first words in b) - d) are proper nouns, the last words refer to sport, b) - d) involve 3rd-singular agreement whereas e) - f) involve 3rd -plural, etc.
Computationally, the compatibility of novel sentences with agreement groups is determined by a mapping process. A novel sentence can be mapped onto an agreement group if the agreement categories of the individual words license a category sequence representing an agreement group. Let us assume the following category assignment for the words in (7) a), our example corpus (letters identify groups, numbers refer to positions (table columns)):

(8)

Adam{b1, c1, d1}	Eve{**b1**, d1}	hates{**b2**, c2, d2}
football{b3, c3, d3, e3, f3}	basketball{**b3**, c3}	hate{e2, f2}

As the boldface items indicate, novel sequence 'Eve hates basketball' can be mapped on group (7) b) because it licenses category sequence *b1b2b3* which in turn symbolises group b).
The mapping algorithm can be extended in such a way that allows for some kind of "guessing" as to the categorical status of unknown words, which broadens the scope of novel sequence processing. For instance, 'People hate football' can be mapped on group e), or f) if 'people' is supposed to have category e1, or f1 (or both).
We remark here that it is also logical to suppose some error-correcting mechanisms to affect natural language acquisition. Consider the examples in (9) from Pinker (1979:240).

(9)

Hottentots must survive
Hottentots must fish
Hottentots eat fish
Hottentots eat rabbits

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

The possible agreement groups are the following:

(10)

a)	b)
Hottentots must survive	*Hottentots must fish*
Hottentots must fish	*Hottentots must survive*
	Hottentots eat fish
c)	d)
Hottentots eat fish	Hottentots eat rabbits
Hottentots must fish	Hottentots eat fish
Hottentots eat rabbits	

Due to the ambiguity of 'fish', groups b) and c) are erroneous as they would license the incorrect 'Hottentots eat survive' and 'Hottentots must rabbits' respectively. A group-correcting mechanism could improve the situation by deleting 'Hottentots eat fish' from b) and 'Hottentots must fish' from c). Alternatively, groups b) and c) could be deleted completely from the learner's memory.

2.1.4. Agreement groups in a distributional context

Agreement groups constitute a kind of distributional approach insofar as the grouping of utterances is determined by the distribution of words they consist of. Distributional methods in linguistics date back at least to Harris (1951). For Harris the distribution of a linguistic item was determined by all the contexts, or "environments" for that particular item. Kiss (1973) proposed a word categorisation model based on cluster analysis which was extended for larger corpora and computational resources by Redington et al. (1998). Finch et al. (1995) adopted a similar method to assign categories to word sequences, i.e. to phrases. Such clustering methods typically operate with "context vectors" as determined by the neighbouring elements of a target item. Mintz (2003) used a different, more direct, formalisation of context. In his work, the immediately preceding and succeeding words provide the context or "frame" for categorising. Mintz employs "frequent frames", i.e. contexts with a frequency larger than an arbitrarily defined threshold. Weisleder and Waxman (2010) consider, besides Mintz's "mid-frames", the usefulness of "end-frames", where the utterance-end marker constitutes an informative element. St. Clair et al. (2010) claim that "flexible frames" exploiting bigram information within frequent frames are more optimal for categorisation than just frequent frames. A kind of framing effect in language acquisition was reported by Cameron-Faulkner et al. (2003) who pointed out that mothers speaking to their children use a rather limited set of item-based phrases, these phrases being framed by their initial words. Such findings were confirmed cross-linguistically by Stoll et al (2009). Below we sketch how agreement groups are related to frames. Cf. Figure 1.

Agreement group:		
1.	Adam hates football	
2.	Adam hates basketball	
3.	Eve hates football	
4.	Adam dislikes football	
5.	Charles hates football	
Frame 1.	Adam X football	X= {hates, dislikes)
Frame 2.	Adam hates X	X= {football, basketball}
Frame 3.	X hates football	X= {Adam, Eve, Charles)

Figure 1. Agreement group as superposition of frames

By substituting words of category X into the appropriate position ("slot") in a frame we get a subset of the agreement group. Frame 1, 'Adam X football', yields sentences 1 and 4, i.e. 'Adam hates football' and 'Adam dislikes football'. Frame 2 licenses sentences 1 and 2, whilst Frame 3 symbolises sentences 1, 3, 5. By forming the union of the utterances licensed by the individual frames we have sentences 1, 2, 3, 4, 5, i.e. the whole group. Thus the agreement group represents a "superposition" of frames.
Wang and Mintz (2010:p. 6) propose that "grammatical relations between words are more consistent in individual frequent frames than in bigrams" and that "words within a frequent frame are especially ''close'' syntactically" (p. 8). This is in accordance with our view that agreement groups represent syntactic (namely, agreement) relations. Bannard and Matthews (2008) suggest that children tend to store word sequences in memory during language acquisition. It could be hypothesized that such word sequences can form the basis of sentence patterns, and that the appropriate grouping of the stored sequences might be a principal element in the emergence of linguistic behaviour. Thus the agreement group method might also be viewed as a kind of model of the organizational processes concerning stored sequences.
Agreement groups can also be regarded as symbolising linguistic patterns representing agreement relations in the generalised sense of Drienkó (2004a, b; 2012b). [7]

3. Results

3.1. Results from previous analyses

Agreement grouping is a straightforward way of structuring a given dataset and the groups obtained can form a basis for processing novel information.

[7] Since Drienkó differentiates between recursive and non-recursive patterns, it is more precise to say that agreement groups symbolise non-recursive agreement relations. 'Recursive' means that certain part(s) of the pattern can be repeated. E.g. 'The boy likes, the girls hate football' can be mapped on a recursive pattern like *(Det N V)^i N* where *(Det N V)* can be repeated arbitrarily many, *i*, times.

Furthermore, the logic behind such a structuring is fairly natural: members of a group belong together because they differ in only some minimal way. This minimal difference, in turn, might be connected to a general capacity of dealing with cognitive/linguistic features. To investigate the applicability of the method to real-life corpora two pilot experiments were carried out on mother-child data. What follows is a brief summary of the results.
The English data for Drienkó (2012a) were gained from the 68 Anne files of the Manchester corpus (Theakston et al., 2001) in the CHILDES database (MacWhinney 2000). The age span was 1;10.7 to 2;9.10. The Hungarian and Spanish utterances for Drienkó (2013a) were also CHILDES data. The Hungarian sessions can be found in the Réger corpus (Réger 1986, Babarczy 2006) collected from Miki, a boy between 1;11 and 2;11. The corpus contains 31 files with the transcriptions of the corresponding audio sessions. The Spanish data were taken from the Montes corpus (Montes 1987, 1992). The child, Koki, is the daughter of an Argentine mother. The recordings were made when the child was 1;7.20 through 2;11.14 old.
The files were converted to simple text format, annotations were removed, together with punctuation symbols. Mother and child utterances were not separated, each dataset was regarded as representing a single 'mother-child language'. Noise was not removed, as it is part of the learning process, i.e. utterances containing 'xxx' were allowed. Similarly, ungrammaticality cannot be separated from learning, so ungrammatical utterances were also included. Only 2-5 word-long utterances were considered.
Each mother-child session, as recorded for the particular corpus, was considered a point in time of linguistic development, a developmental stage. Each point represented the linguistic knowledge acquired up to that point. It was tested to what extent the agreement groups at a given stage can account for the utterances of the immediately following session. Agreement groups were extracted from the body of all the utterances, meeting the above criteria, encountered up to the test stage. Each utterance had its own group. It was investigated to what extent the utterances of the next session could be 'mapped' onto the already existing agreement groups.

The experiments revealed the following facts:

1. At every developmental stage there are novel utterances compatible with some agreement group. Additionally, extended mapping (guessing of categories) may improve processing.
2. Agreement categories (as modelled by our hypothetical table columns) may relate quite naturally to notions like "lexical category", and "semantic category".
3. Traces of grammatical agreement can be found in agreement groups.

In order to prime the understanding of what is to be said about 'coverage' in Section 4, we restate our mapping results here, i.e. the findings corroborating Fact 1 above.
The results concerning the proportion of English utterances that were compatible with at least one agreement group are visualised for each

developmental stage in Figure 2. The range of mapped novel utterances is 6% - 10.3%. It is 6% - 8.9% for the first part of the diagram (for the first 14 sessions) and 7.3% - 10.3% for the later developmental stages, which suggests a slight increase. The figure also displays that some 19% - 41% of the total amount of utterances, both novel and non-novel, were compatible with at least one group. For non-novel utterances compatibility actually means that the very same utterance had already been heard by the child, so the utterance has its own group and may be a member in others, which makes mapping trivial.

Figure 2. Mapping results from Drienkó (2012a) for the Anne data, Manchester corpus. Child's age: 1;10.7 to 2;9.10. Maximum training set size:

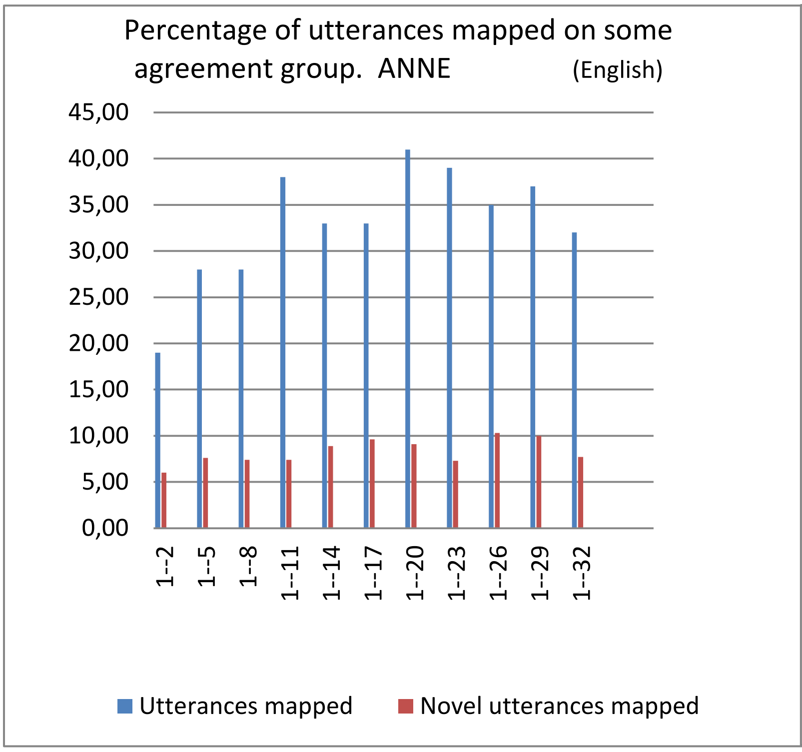


Figure 3. Mapping results from Drienkó (2013a) for the Koki data, Montes corpus. Child's age: 1;7.20 to 2;11.14. Max. training set size: 2150

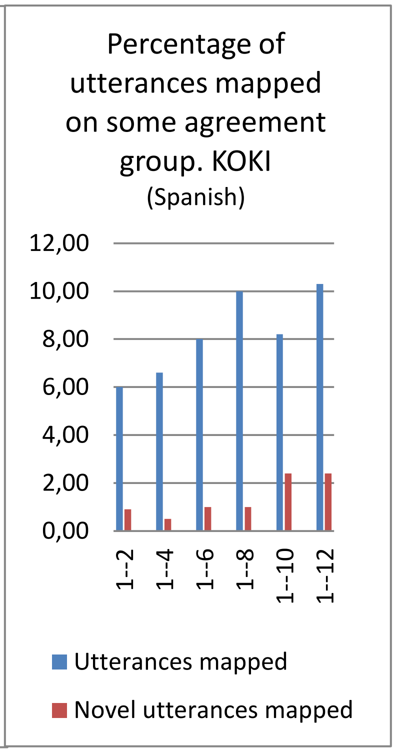


Figure 3 shows the mapping details for the Spanish case. The range of mapped novel utterances is 0.5% - 2.4%. A slight increase is observable, from 0.5% - 1% for the first three stages to 1% - 2.4% for the later stages. Of all the test file utterances, both novel and non-novel, some 6% - 10.3% were compatible with at least one agreement group.

The diagram from the Hungarian analysis is displayed as Figure 4. The range of mapped novel utterances is 1.2% - 7.7%. There is a clear increase, from 1.2% - 4% for the first part of the diagram (the first 7 sessions) to 3.2% - 7.7% for the later developmental stages. Of all the test file utterances, both novel and non-novel, some 5% - 18.4% were compatible with at least one group.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

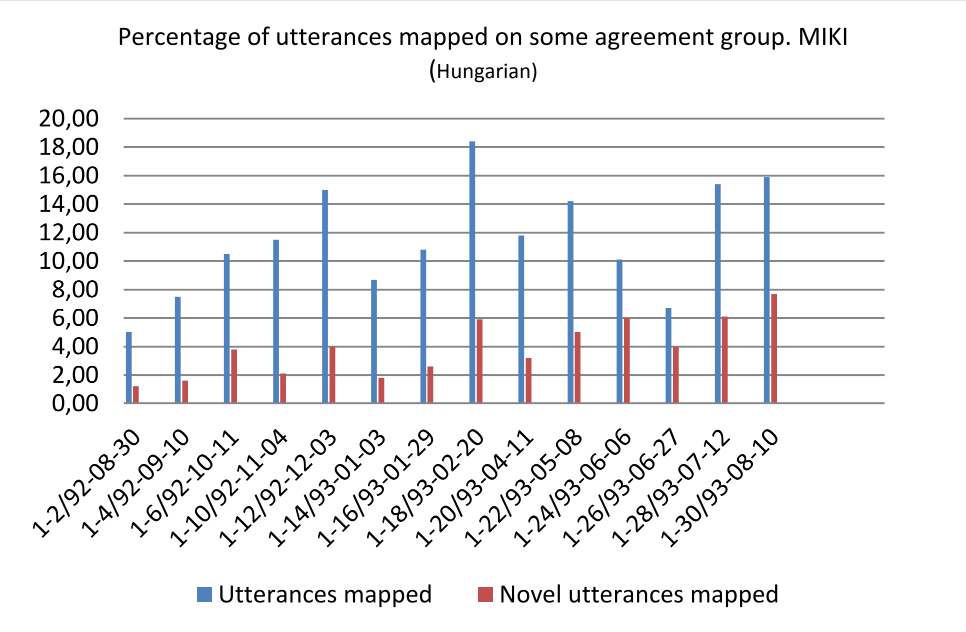


Figure 4. Mapping results from Drienkó (2013a) for the Miki data, Réger corpus. Age: 1:11 to 2:11. Max. training set size: 7122 utterance types.

If we assume the learner to "know" or rather "expect" that any new word in a grammatical utterance must observe the agreement requirements of some agreement group, we can extend the mapping mechanism to include unfamiliar words. Figure 5 contrasts the two ways of mapping for the English data. The first 64 Anne files, 1a through 32b, provided the agreement groups, whereas files 33a-34b were used for testing. As expected, extended mapping resulted in higher compatibility with test utterances. The maximum of novel utterances processed became 12.2% from 8.9%, while the overall maximum increased to 42% from 38%.

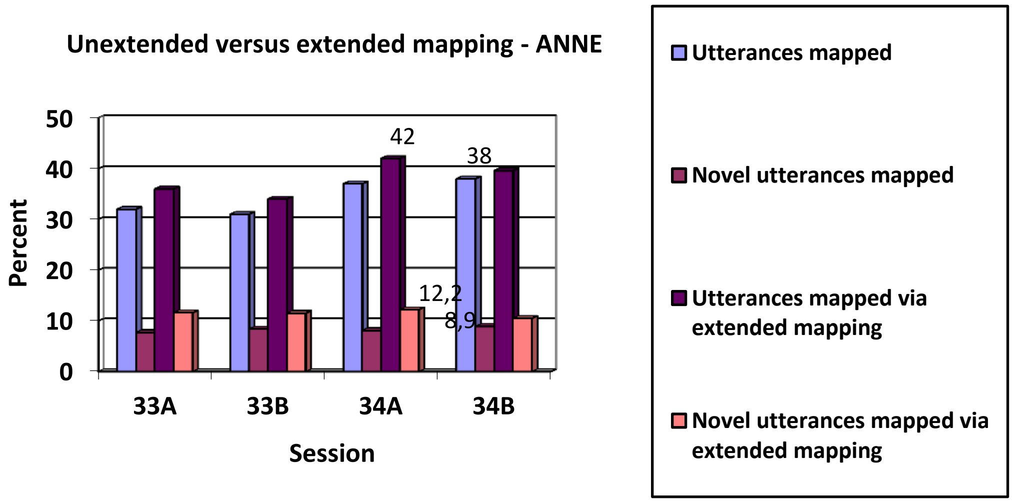


Figure 5. Drienkó (2012a)'s English results for contrasting unextended and extended (category guessing) mapping

The Spanish and Hungarian results suggest a more significant role for extended mapping. The data obtained via the two distinct mapping mechanisms are juxtaposed in Figures 6 and 7 for the Spanish and the Hungarian utterances, respectively. As visualized in Figure 6 for the Spanish files, the maximum of mapped utterances rose from 10.3% to 25.5% due to the extension of the mapping mechanism, whereas the maximum of mapped novel utterances rose very noticeably from 2.4% to 18.5%. For the Hungarian data the maximum of mapped utterances rose from 18.4% to 36%, and the 7.7% maximum of mapped novel utterances became 25.2% in the extended case. Cf. Figure 7.

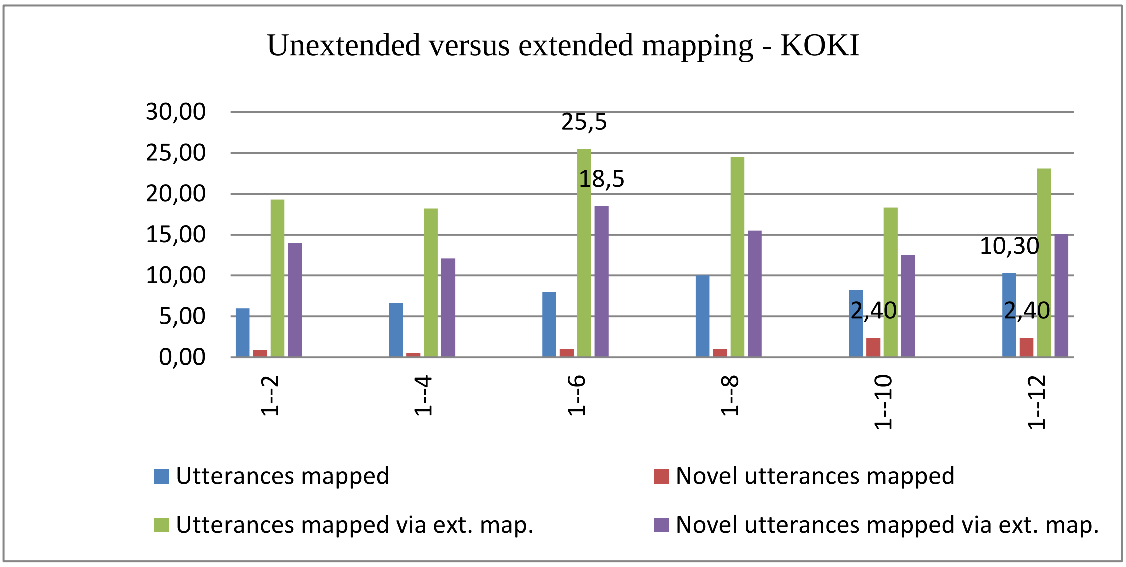


Figure 6. Drienkó's (2013a) Spanish results for contrasting unextended and extended (category guessing) mapping.

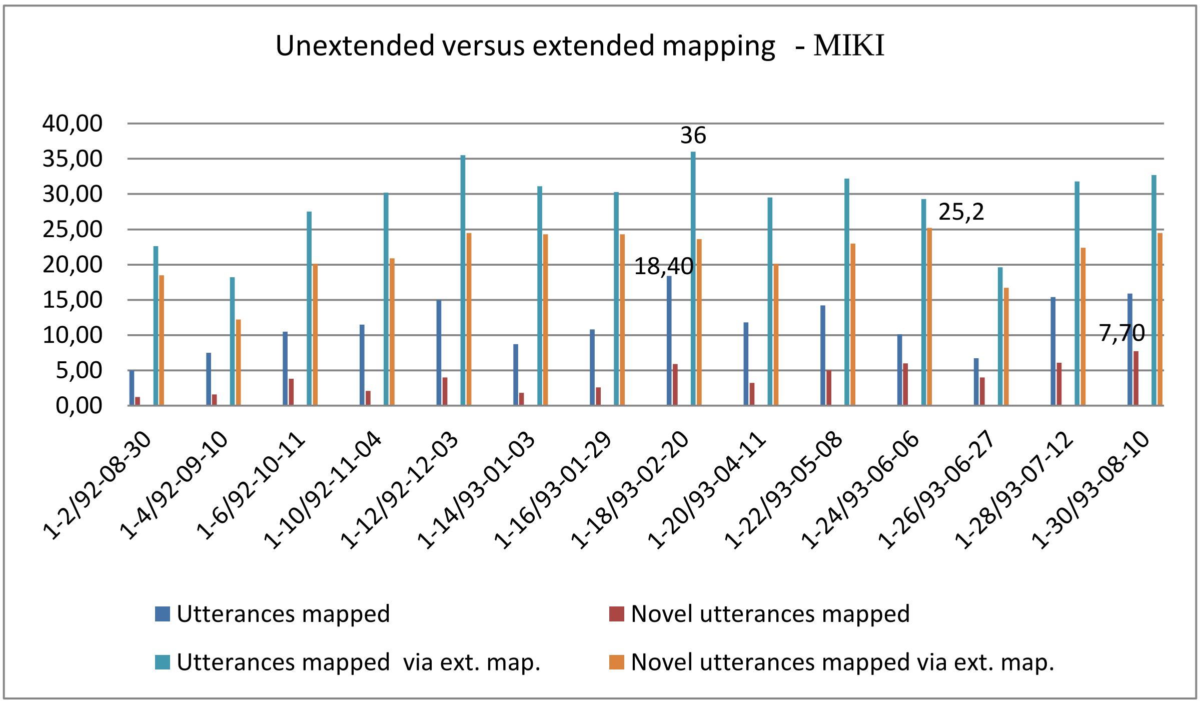


Figure 7. Drienkó's (2013a) Hungarian results for contrasting unextended and extended (category guessing) mapping

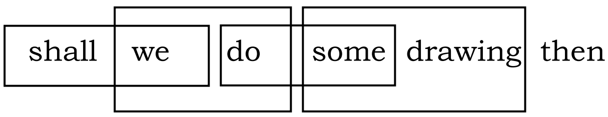
3.2. Agreement groups coverage

The agreement groups in the previous analyses consisted of utterances whose length did not exceed five words. In the general case, however, natural language utterances can be considerably longer – infinite, in theory. Thus agreement groups analysis seems to be inapplicable for utterances longer than an arbitrarily chosen upper bound. Nevertheless, if we view utterances as possibly composed of smaller linguistic fragments, agreement groups may readily offer themselves for handling such fragments. Accordingly, in our next experiment we do not expect a group to account for a whole utterance in each case. Instead, we examine in what measure an utterance can be 'covered' by smaller fragments corresponding to some agreement group(s).

To give an immediate introduction to what we mean by 'coverage', we consider the example sentence '*shall we do some drawing then*'. All the possible – non-discontinuous – fragments of this sentence are shown under (11). All fragments consist of two through five words – cf. our arbitrary restriction on sequence length. We check agreement-mappability for each fragment separately, i.e. as if they were separate utterances. The boldface elements indicate that four fragments, 'shall we', 'we do', 'do some', and 'some drawing' can be mapped onto at least one agreement group. This result is visualized in (12).

(11)
shall we :shall we do :shall we do some :shall we do some drawing :**we do** :we do some :we do some drawing :we do some drawing then :**do some** :do some drawing : do some drawing then :**some drawing** :some drawing then :drawing then :

(12)



Coverage for a sentence, then, is calculated in a fashion sketched by (13): if an utterance position (measured in words) is covered by a legal fragment (i.e. a fragment compatible with some agreement group), that position scores 1. If there is no legal fragment to cover a position, the position scores 0. For the situation represented by (13) the coverage is 83% because five utterance positions of all the six are covered by the four fragments 'shall we', 'we do', 'do some', and 'some drawing'. Note that with this definition of 'coverage' it is irrelevant how many fragments are involved in the covering of a certain position, i.e. how many times the position is covered.

(13)

1	2	3	4	5	6
shall	we	do	some	drawing	then
1	1	1	1	1	0

Coverage$_1$: five words of six = 5/6 = 83%

Since agreement groups have a potential for processing novel sequences it may be instructive to test that potential in the coverage case, too, i.e. on utterances regarded as fragmentary. In other words, we can examine the degree of novelty detectable in fragments. Utterance 'shall we' in (11), for instance, is a legal fragment since it is compatible with some group. However, a closer look reveals that this compatibility is trivial because 'shall we' is an explicit utterance of the training corpus so it is compatible with any group containing it. Thus (14) is a more finely drawn picture of the coverage situation than (12). The discontinuous box indicates that 'shall we' is not novel. Continuous lines, on the other hand, are for novel fragments.

(14)

shall we do some drawing then

Evidently, the coverage result for '*shall we do some drawing then*' will be different when only novel fragments are taken into account. Cf. (15).

(15)

1	2	3	4	5	6
shall	we	do	some	drawing	then
0	1	1	1	1	0

Coverage$_2$: four words of six = 4/6 = 66%. (Novel fragments only.)

Table 2 illustrates how individual groups can provide the fragments employed by the coverage mechanism. The fragment 'shall we', for example, is compatible with the 'shall I' group. 'shall we' is actually a member of the group so compatibility is trivial, as noted earlier. 'we do', on the other hand, is a novel fragment and can be mapped on the 'I xxx' group. [8]

[8] The 'I xxx' group may be seen as rather mixed to serve as a fair basis for providing legal fragments, especially if 'legal' is understood as 'grammatical'. With this problem, we would like to refer the reader to Drienkó (2014)'s comments on the precision of his method. Briefly, 'heterogeneous' groups may correspond to some transitional stage of development. Such groups can be augmented over time by an appropriate error-correction mechanism. Here we add that the emergence of just a few erroneous fragments in the covering process might not necessarily be critical. The omission of the 'we do' fragment, for instance, would not influence coverage significantly in the example. For the non-novel case coverage would actually remain the same. Cf. (12) and (13), as well as the remarks concerning (27) in the Discussion.

Table 2.

Four agreement groups licensing fragments 'shall we', 'we do', 'do some', and 'some drawing'

SHALL WE									
shall I	**SHALL WE**	don't I	aren't I	willn't I	have I	do I	did I	haven't I	and I

WE DO						
I xxx	tip xxx	head xxx	I know	want xxx	I comfy	I writing
we've xxx	fit xxx	I carry	is xxx	some xxx	and xxx	lorry xxx
water xxx	baby's xxx	I see	oh xxx	that's xxx	no xxx	I sleeping
I up	that xxx	where xxx	get xxx	more xxx	I cry	baby xxx
horsie xxx	I just	Anne's xxx	I can't	I think	like xxx	I did
one xxx	good xxx	hello xxx	I **do**	I stand	kiss xxx	Spot xxx
snake xxx	I don't	I can	I will	I sleep	I home	okay xxx
I am	my xxx	what xxx	I'll xxx	I snoozing	yeah xxx	yes xxx
phone xxx	I bet	I coming	I not	I get	I asleep	I stuck
Mummy xxx	who's xxx	I promise	got xxx	a xxx	they xxx	he's xxx
it's xxx	you xxx	I would	I say	I have	don't xxx	his xxx
have xxx	daddy xxx	happy xxx	I want	I haven't	need xxx	I like
there xxx	can xxx	Anne xxx	I does	come xxx	elephant xxx	Daddy xxx
here xxx	just xxx	I had	I got	I willn't	**we** xxx	I shouldn't
I didn't	golden xxx	this xxx	I lost	down xxx	I landed	butter xxx
I hungry	I gathered	I dripped	I done	I weren't	think xxx	I doing
I lots	I suppose	I need	in xxx	I noticed	now xxx	I wonder
penguin xxx	strawberry xxx	bye bye xxx	he xxx	boy xxx	not xxx	I going
two xxx						

DO SOME									
got it	xxx it	like it	sweep it	eat it	look it	dropped it	drop it	open it	perhaps it
that's it	lost it	wipe it	tip it	**do** it	push it	pushed it	move it	done it	got spade
read it	isn't it	bite it	carry it	get it	show it	find it	found it	spill it	chewing it
throw it	put it	leave it	have it	see it	splash it	reach it	hold it	cut it	painting it
stop it	mend it	brush it	stand it	stroke it	sing it	tore it	smack it	caught it	got **some**
is it	want it	tear it	got her	eating it	rub it	where's it	did it	doesn't it	broken it
has it	got one	got xxx	got cow	break it	got over	need it	spilling it	park it	missed it
will it	got to	wet it	got pram	got him	willn't it	hasn't it	can't it	didn't it	got cheese
not it	was it	rolling it	mix it	fly it	make it	got presents	got ticket	bumped it	because it

SOME DRAWING					
my juice	want juice	of juice	drink juice	oh juice	Anne's juice
more juice	my cover	my feet	your juice	my curtain	my curtains
bottle juice	my finger	my me	xxx juice	my drink	my xxx
my Mummy	my phone	my side	my shops	my sweet	my toes
my eyes	my legs	my bag	my fridge	my cot	my can't
my toys	my game	my reindeer	my driver	my umbrella	my tummy
my back	my grapes	my name	my hand	my fault	**some** juice
my favorite	my book	my nose	my leg	my coat	my special
and juice	my cheeks	my milk	my **drawing**		

3.2.1. Coverage results

The results of a coverage analysis were first reported in Drienkó (2013b). The method was applied to the Anne data of the Manchester corpus (Theakston et al. 2001), CHILDES database (MacWhinney 2000). Table 3 lists the basic input parameters for the analysis.

Table 3.
Dataset details for the coverage experiment

Training corpus data			Test file data	
File ID	Group space (utterance types)	Number of words (types)	File ID	File size (utterance types)
1a-32b	17260	2505	33a	466

The first 64 Anne files, 1a through 32b, provided the agreement groups, whereas the 466 utterance types[9] in File 33a were used for testing. We calculated coverage for each individual utterance type and divided the sum of the individual coverage values by 466, the number of utterance types, for an average, as shown in Figure 8. The figure also reveals that the average coverage, c_a, was equal to 78 percent. When only novel fragments were considered the average coverage was reduced to c_a = 49%.

Average coverage

Overall:

$$c_a = \left(\sum_{i=1}^{n} c_i \right)/n = 364.54/466 = 78\%$$

(c_i : coverage for sentence$_i$, n=466)

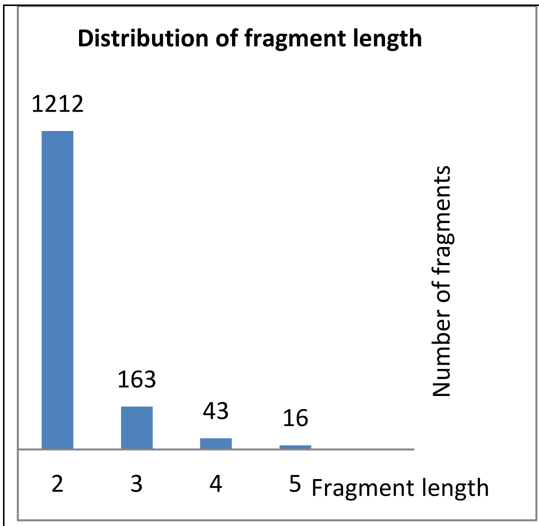
Novel fragments only:

$$c_a = 229.02/466 = 49\%.$$

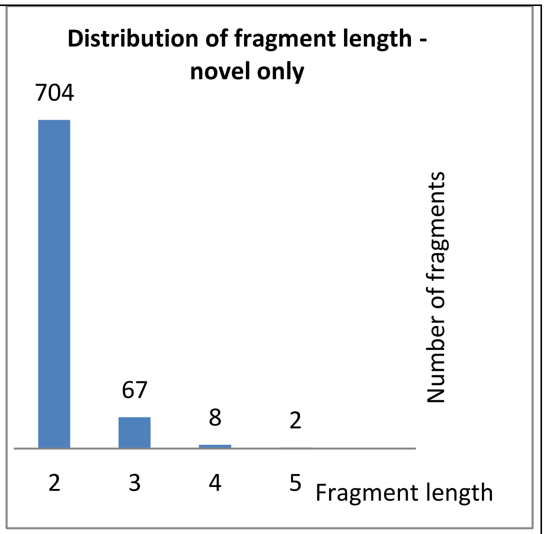
Figure 8. The results of the coverage experiment.

Given that a fragment could be minimally two, maximally five words long, it may be of some interest to see how many words a legal fragment consisted of on average. The distribution diagrams of Figure 9 indicate that most fragments were two-word-long. The average length was between two and three words: 2.21 words when all legal fragments were considered, and 2.11 words for the novel fragments only case. This may be in accord with our intuition that the shorter utterances a group contains, the more utterances it has. More utterances, in turn,
may mean a greater probability of exactly matching a fragment. More utterances in a group may also mean more 'generative' power, i.e. a larger capacity for novelty, hence a larger degree of compatibility for novel fragments. This is supported by the fact that the average length for novel fragments was slightly less (2.11 words) than in the overall case (2.21 words).

[9] One-word utterances were not included.



Average fragment length: 2.21 words



Average fragment length: 2.11 words

Figure 9. Fragment size statistics both for the general, and for the 'novel only' case.

3.2.2. Discontinuous coverage

The previous experiment treated fragments as undividable, i.e. non-discontinuous, sequences. On the other hand, natural language utterances do exhibit dependencies between elements (words) thought to belong to the same linguistic unit (constituent) when separated by other elements (words). Consequently, it would be logical to examine how coverage can be influenced by assuming a possible discontinuity within fragments. Our next experiment included discontinuous fragments.

As comparing the examples in (11) and (16) reveals, there are much more fragment candidates for *shall we do some drawing then* in (16) where we allow words that are not adjacent in the original utterance to possibly belong to the same fragment. This would predict that taking dependencies between separated words into account results in higher coverage.

(16)

drawing then :some then :some drawing :some drawing then :do then :do drawing :do drawing then :do some :do some then :do some drawing :do some drawing then :we then :we drawing :we drawing then :we some :we some then :we some drawing :we some drawing then :we do :we do then :we do drawing :we do drawing then :we do some :we do some then :we do some drawing :we do some drawing then :shall then :shall drawing :shall drawing then :shall some :shall some then :shall some drawing :shall some drawing then :shall do :shall do then :shall do drawing :shall do drawing then :shall do some :shall do some then :shall do some drawing :shall do some drawing then :shall we :shall we then :shall we drawing :shall we drawing then :shall we some :shall we some then :shall we some drawing :shall we some drawing then :shall we do :shall we do then :shall we do drawing :shall we do drawing then :shall we do some :shall we do some then :shall we do some drawing :

Indeed, Table 4 shows that besides the fragments found in the non-discontinuous case (printed bold), there are further fragments that can contribute to the coverage of the utterance and due to the fragments containing *then*, the whole utterance is covered.

Table 4.
Coverage (6 words of 6): 6/6=100%

Shall	*we*	*do*	*some*	*drawing*	*then*
			some		then
			some	**drawing**	
		do			then
		do	**some**		
	we				then
	we			drawing	
	we		some		
	we	**do**			
shall					then
shall			some		
shall		do			
shall	**we**				

Again, the first 64 Anne files, 1a through 32b, provided the agreement groups, whereas the 466 utterance types in File 33a were used for testing. Figure 10 displays the result. In accord with intuition, average coverage, c_a, became higher, 83 percent. Compared to the 78 percent value of the non-discontinuous case this amounts to some five percent increase.

Average coverage – discontinuous case

$$c_a = \left(\sum_{i=1}^{n} c_i\right)/n = 385.64 \;/\; 466 \;=\; 83\%$$

(c_i : coverage for sentence$_i$, n=466)

Figure 10. Coverage results for the discontinuous case

3.3. *Theoretical implications: linguistic processing and levels of generalisation*

The above series of experiments reveals that agreement grouping may provide some insight with respect to how language builds up in time, i.e. what kind of linguistic representation may be necessary, what types of cognitive mechanisms may underlie linguistic processing: Below we show, in the context of linguistic generalisation, how our observations can yield theoretical, even formalised, results. We will delineate three basic levels of generalisation and characterise them in terms of the "abstract" structures involved, and the computational requirements they impose on the processing mechanism.

3.3.1. Three levels of generalisation

A key question in language acquisition is how children learn to abstract away from the actual linguistic material they hear, to extend their knowledge to process novel utterances. Obviously, at the same time, children's generalisations must be constrained so that they can produce only "grammatical" utterances. In the present section we list the generalisation mechanisms that a model, based on two cognitive capacities, grouping and finding smaller fragments in larger ones (i.e. coverage), can offer. The three levels of generalisation we will present constitute a "continuum" in that each subsequent level builds on the abstract representations of the previous one(s) and requires more computational complexity.

Level 1

The first level of generalisation/abstraction was actually introduced in Section 2.2 where we saw how the tabular representation of agreement groups can be used for processing novel utterances. Also, we drew a parallel between agreement groups and linguistic patterns in Section 2.4. The computational requirement for this abstraction level consists in finding a sequence of agreement categories which is compatible with some agreement group.

Level 2

To reach the second level of generalisation, we need the notion of "coverage structure". Note that we cannot use boxes like e.g. in (12) to depict how possibly discontinuous fragments contribute to the coverage of an utterance. A tabular representation, like Table 4, on the other hand, is a more appropriate visualisation of what we will call the "coverage structure" of an utterance. Furthermore, coverage structure may form a basis for "constituent structure" if we know which fragments to choose and how to combine them. For instance, Table 5 shows a subset of the fragments in Table 4 arranged in such a way that we get a quasi constituent-tree by iteratively enclosing the content of the row in brackets, moving words (and brackets) downwards as the arrows indicate, and coalescing identical words. The "derivation" is shown in (17).

Table 5
Quasi constituent-tree

Shall	*we*	*do*	*some*	*drawing*	*then*
			some↓	drawing↓	
		do↓	some↓	↓	
	we↓	do↓	↓	↓	
shall↓	↓	↓	↓	↓	then↓

(17) some drawing → (some drawing) → do (some drawing) →
(do (some drawing)) → we (do (some drawing)) → (we (do (some drawing))) →

→ shall (we (do (some drawing))) then→ (shall (we (do (some drawing))) then)

Thus, besides representing discontinuous dependencies, coverage structure may help assign constituent structure to utterances. What is more important for the present discussion is that coverage structure enables cross-fragment generalisation, i.e. "abstraction" across agreement groups participating in the coverage of an utterance. Recall that fragments are mapped onto agreement groups and words receive agreement categories depending on their positions within utterances. In (18), e.g., we can see the fragments of Table 4 translated into sequences of agreement categories.[10] For example, fragment *some then* can be mapped onto group 292, where *some* can appear in the first position (hence its category 292_1).

(18)

some then	292_1 292_2	some drawing	304_1 304_2	do then	2_1 2_2
do some	379_1 379_2	we then	681_1 681_2	we drawing	5547_1 5547_2
we some	551_1 551_2	we do	55_1 55_2	shall then	12969_1 12969_2
shall some	16636_1 16636_2	shall do	16278_1 16278_2	shall we	136_1 136_2

We assume that language learners memorise utterances and make generalisations on the basis of such stored utterances. For short utterances agreement grouping works well. However the one-word similarity criterion would require an unrealistically large corpus of stored sequences for longer utterances. On the other hand, if we suppose that learners memorise utterances together with their coverage structures then we can reach a next (second) level of generalisation. Coverage structure provides immediate clues as to which fragments can be combined with which others within a grammatical utterance. The fact that fragments can be identified with groups is the key point here and is visualised in Table 6, where fragments are represented as category sequences.

Table 6
Fragments

Shall	*we*	*do*	*some*	*drawing*	*then*
			292_1		292_2
			304_1	304_2	
		2_1			2_2
		379_1	379_2		
	681_1				681_2
	5547_1			5547_2	
	551_1		551_2		
	55_1	55_2			
12969_1					12969_2
16636_1			16636_2		
16278_1		16278_2			
136_1	136_2				

[10] Note that fragments may be mapped onto several groups, which possibly means several category assignments for a given fragment. Here this would simply mean more rows in the corresponding table, and consequently greater generalisation potential.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

Coverage structure, then, is not just a snapshot of the dependencies and structural relations of a given utterance: it is a pattern or "schema" which enables the processing of a whole class of linguistic sequences. More formally, a schema is a coverage structure where fragments are represented with categories (cf. Table 6). An utterance *U* can be mapped onto schema *Sc* if there is a subset Sc_k of the fragments in *Sc* such that *U* can be covered by Sc_k, i.e. by the groups corresponding to the fragments of Sc_k. For instance, alongside with many grammatical and ungrammatical sentences *Have I got some juice then* is compatible with the schema of Table 6. Categories printed bold in Table 7 represent the subset of groups which can be used in covering the utterance with the fragments in (19) as Table 8 shows.

(19)

some juice	304_1 304_2	got some	379_1 379_2
I got	55_1 55_2	have then	12969_1 12969_2

Table 7

Shall	*we*	*do*	*some*	*drawing*	*then*
			292_1		292_2
			304_1	**304_2**	
		2_1			2_2
		379_1	**379_2**		
	681_1				681_2
	5547_1			5547_2	
	551_1		551_2		
	55_1	**55_2**			
12969_1					**12969_2**
16636_1			16636_2		
16278_1		16278_2			
136_1	136_2				

Table 8

Have	*I*	*got*	*some*	*juice*	*then*
			some	juice	
		got	some		
	I	got			
have					then

Note that we get a simplified "parse" for an utterance if we consider the columns of schemas to be symbolising "metacategories" associated with sets of categories. On such a view schemas are just sequences of metacategories and mapping onto a schema simply means category matching. For example, the first column of Table 6 consists of categories 12969_1, 16636_1, 16278_1, and 136_1, so it is a "12969_1 16636_1 16278_1 136_1" metacategory. The schema of Table 6 is compressed into a metacategory sequence as Table 9. Similarly, we can interpret utterances as sequences of metacategories since a word can belong to several agreement categories.

Table 10 presents *Have I got some juice then* as a sequence of metacategories.[11]

Table 9

Shall	*we*	*do*	*some*	*drawing*	*then*
12969_1	681_1	2_1	**292_1**	**304_2**	**292_2**
16636_1	5547_1	**379_1**	**304_1**	5547_2	**2_2**
16278_1	551_1	**55_2**	**379_2**		**681_2**
136_1	**55_1**	16278_2	**551_2**		**12969_2**
	136_2		**16636_2**		

Table 10

Have	*I*	*got*	*some*	*juice*	*then*
12969_1	**55_1**	**55_2**	**292_1**	**304_2**	**292_2**
…	…	**379_1**	**304_1**	…	**2_2**
		…	**379_2**		**681_2**
			551_2		**12969_2**
			16636_2		…
			…		

If we define matching of metacategories as having at least one element (agreement category) in common, the metacategory sequence for *Have I got some juice then* matches that of the schema of Table 9, which in turn can be interpreted as *Have I got some juice then* being mappable onto that schema. Cf. the boldface items. The advantage of this metacategory sequence interpretation might be that it allows a simple model of abstraction: word sequences of encountered utterances become category sequences enabling the processing of novel utterances. Ontogenetically, metacategories could be the precursors for more abstract syntactic categories like noun, verb, etc.
Basically, the computational requirement for Level 2 is to find a subset of fragments (groups) in a schema that can cover the input utterance. For the metacategories case the task is to find a schema matching the category sequence of the utterance in question. Schemas are "composed" of groups, the primary generalisation engines of Level 1.

Level 3
Once again, the above level of "abstraction" is a consequence of the fact that fragments in a schema represent groups. A further level of generalisation, "cross-schematic abstraction", can be reached by discarding the idea that a fragment-type (i.e. group) can only contribute to the construction of the particular utterance-type in which it was detected and memorised by the learner. If we regard a coverage structure comprising k fragments as guidelines for how any two of the k fragments (i.e. any two groups of the k) can be combined we obtain $k(k-1)/2$ partial constraints on coverage structure in general. Putting together such partial "combinability" constraints provided by all schemas yields a pool of constraints responsible

[11] Recall that words can possibly have much more categories than those indicated in the table.

for the coverage structure of principally any utterance. Combinability constraints (or "rules") can be extracted automatically from schemas by simply recording the relative order of the elements of the two fragments in question. More formally put, a combinability constraint *CC* for fragments F_1 and F_2 in schema *Sc* is a sequence of agreement categories $c_1...c_n$, where each c_i ($1 \leq i \leq n$) occupies a unique position in *Sc*, each c_i belongs to either fragment F_1 or F_2, and if c_i precedes c_j in *Sc* then c_i precedes c_j in *CC*. If c_i and c_j appear in the same column of *Sc* their order is immaterial in *CC*.[12] Agreement categories appearing in different columns of *Sc* are separated by commas. For instance, based on the schema in Table 6, the combinability constraints for some groups are given in (20). Interestingly, combinability constraints can typify syntactic operations/phenomena. *CC(55, 304)*, the combinability constraint for groups 55 and 304, licenses a kind of juxtaposition or concatenation of strings coming from the respective groups, but in a broad sense since fragments corresponding to other groups can intervene if we step out of schema Table 6. The constraint in (20) b) effects the embedding of group 379 fragments in fragments from group 681, while constraint *CC(2, 5547)* results in a crossing dependency effect. (20) d) can be seen as a pattern for substituting a group 304 fragment for category 292_1. Note that since the relative position of any two fragments can vary in schemas, there can be several combinability constraints for the same pair of fragments (groups).[13]

(20)

a)	CC(55, 304)	55_1, 55_2, 304_1, 304_2,	"concatenation"
b)	CC(379, 681)	681_1, 379_1, 379_2, 681_2,	"embedding"
c)	CC(2, 5547)	5547_1, 2_1, 5547_2, 2_2,	"crossing dependency"
d)	CC(292, 304)	292_1 304_1, 304_2, 292_2,	"substitution"

A "coverage parsing" algorithm is outlined as (21).

(21)

1. Given: input utterance *U*
2. Cover the utterance with possible fragments, i.e. set up coverage structure *CS* for *U*
3. Find a coverage structure CS_k (subset of *CS*) consisting of *k* fragments arranged in *k* rows in *CS* such that any two fragments, F_i and F_j ($1 \leq i, j \leq k, i \neq j$) in CS_k be combinable, i.e. for any pair of fragments there exist a combinability constraint *CC(i,j)* or *CC(j,i)*.
4. *U* is grammatical if CS_k exists (i.e. if every word position in *U* is covered by CS_k)

12 For simplicity the agreement category represented by a smaller number will be written first.
13 Thus it might be more precise to write $CC(i,j)_k$, where the index k would differentiate constraints for the same pair of groups.

In case the notion of grammaticality does not seem theoretically clear-cut enough, it is possible to relax our grammaticality criterion and employ a milder notion of acceptability instead: one might, e.g., require the existence of combinability constraints for only some part of CS_k depending on the degree of acceptability one wishes to tolerate.

The computational requirement for Level 3 is to find a subset of fragments (groups) with valid combinability constraints that can cover the input utterance. CCs are abstracted from schemas, the primary generalisation objects of Level 2.

Our definitions concerning CCs and the "parsing" algorithm (21) will be made more general in Section 6.3.

4. Discussion

4.1. Agreement groups for usage-based syntactic processing

The basic findings of the coverage experiments are listed under points 1-5 below.

1. Longer utterances can be covered by shorter ones (fragments) that are compatible with agreement groups.
2. Agreement groups can account for a large part of a covered utterance (78% on average, in our non-discontinuous experiment).
3. There is a considerable degree of novelty in fragments.
4. Agreement groups can account for the novelty in fragments (on average, 49% of an utterance was covered by novel fragments, i.e. by word sequences which were not explicitly present in the training corpus).
5. Allowing discontinuous fragments enhances coverage.

These findings, taken together with previous results on agreement groups, may also have some theoretical relevance as to the understanding of the acquisition and processing of linguistic material, as was specifically shown in terms of generalisation levels. Roughly, the data seem to suggest two major levels of processing. Firstly, there is a more elementary, 'group-level' mechanism responsible for the linguistic appropriateness of the more elementary utterances belonging to this level. Secondly, these elementary utterances may either be standalone utterances in their own right, or they may qualify as building blocks, i.e. fragments, for longer, more complex utterances. More complex sentence building, the combination or integration of fragments, might then be done by the processing mechanisms of the more complex 'covering level'. It is also possible to demarcate sublevels for both major levels. For instance, above we hypothesised that the coverage level may consist of Level 2 and Level 3, characterised by schemas and combinability constraints, respectively. Similarly, group-level processing can also be graded, in accordance with how many "slots" are allowed in a particular frame, ranging from zero to the number of all the word positions in the group (set arbitrarily to 5 in our experiments). Zero slots means no generalisation at all. This is the case when each agreement category – each column in the corresponding table – has only one word, i.e. we have a "one-member group" consisting of a single stored utterance. The other extreme is maximal generalisation when each category has more than one word. Such a division of processing levels might be consonant with the three usage-based options for children to produce an utterance as described in Tomasello (2003, pp. 308-309). The first option, just producing an utterance as it was heard – *There-ya-go*, for example – corresponds to zero generalisation. The second option is retrieving a stored utterance and perform some elementary operations on it. These operations can be i) fitting a new constituent into a slot, ii) adding a new constituent onto the beginning or the end of a retrieved utterance, iii) insert a new constituent into the middle of a retrieved utterance. While slot-filling is typically a group-level operation in our framework, adding and inserting new elements involves the coverage levels. The third option, "combining constituent schemas" may point in the direction of combinability constraints, associated with our highest level of generalisation.

The elementary operations for ii) and iii) above stem from Lieven et al. (2003) who employ five basic operations – substitute, add on, drop, insert, and rearrange – to derive a closest match for a novel utterance. The number of operations is reduced in Dąbrowska and Lieven (2005) who employ only two: juxtaposition and superimposition.

The positioning of fragments relative to each other might be characterised in terms of similar basic operations in the agreement groups model, as well. The non-discontinuous covering algorithm that we employed in our first coverage experiment was able to detect only two fragment combination types, as apparent from (12). We will refer to these as 'contiguous' and 'intersecting', respectively. Cf. (22) and (23). They suppose a non-discontinuous succession of fragments in time, i.e. no word from another fragment comes between any two words of any fragment. Non-discontinuity does not preclude the sharing of elements in ending/beginning positions, as (23) exemplifies. The contiguous example of (22) can be seen as simply concatenating or "juxtaposing" fragments, while the intersecting fragments of (23) might exemplify a primitive way of superimposition – which might mean substitution, as well: e.g. *we do* is substituted for *we*. (22) can also be seen as an example of add on: either *shall we* is added on *do some* or vice versa. Our one-word-difference criterion on agreement groups is closely related to the notion of substitution: any non-base member of a group can actually be seen as a result of applying a single substitute operation to the base utterance, while novel utterances may alternatively be regarded as multiple-operation substitute sequences.

(22)

Contiguous fragments ('concatenation', "juxtaposition")

Fragment1 Fragment2

$\boxed{\text{shall we}}\boxed{\text{do some}}$

(23)

Intersecting fragments ("superimposition")

Fragment1 Fragment2

shall we do

Our results showed that just these two types of non-discontinuous fragment combinations were sufficient to cover 78 percent of mother-child data. Allowing discontinuous fragments improved coverage by a modest five percent. The increase is not much. However, bearing in mind that mother-child discourse is actually a simplified way of using language, one should not expect mother-child data to include a significant number of complicated linguistic constructions whose analysis would require the more powerful apparatus of discontinuous processing. That said, languages do have complicated constructions like embeddings or crossing dependencies. These constructions cannot be processed with just contiguous, and intersecting fragments, since they involve discontinuity. The hypothetical examples under (24) reflect a kind of embedding effect. The fragment *not very* in (24) a) is embedded in fragment *a nice girl* without the sharing of any words – a kind of INSERT operation. On the other hand, (24) b) shows a sort of substitution effect owing to the fact that *nice* belongs to both fragments: *nice* is replaced with *not very nice*.

(24)

EMBEDDING FRAGMENTS

a) Fragment1: **a nice girl** Fragment2: not very

a not very **nice girl**

b) Fragment1: **a nice girl** Fragment2: not very nice

a not very **nice girl** ('substitution': nice → not very nice)

The nested dependency construction of (25) demonstrates the possibility of multiple embeddings for fragments. Table 11, the coverage structure for *The rat the cat the dogs hate chased eats the cheese*, also reveals that within-fragment, "local", agreement relations remain valid for the whole utterance. The nouns' person and number values agree with those of the corresponding verbs: the rat eats whereas the dogs hate.

(25)

Multiple embedding (Nested dependency)

Fragment1: **The rat eats the cheese** Fragment2: *The cat chased*

Fragment3: The dogs hate

(**The rat** (*The cat* (The dogs hate) *chased*) **eats the cheese**)

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

Table 11
The coverage structure for 'The rat the cat the dogs hate chased eats the cheese'

The	*rat*	*the*	*cat*	*the*	*dogs*	*hate*	*chased*	*eats*	*the*	*cheese*
the	rat							eats	the	cheese
		the	cat				chased			
				the	dogs	hate				

The fact that such constructions are virtually limited to two embeddings, may be due to differences in generalisation levels. While constructions with just one or two embedded clauses can be stored as schemas – Level 2 generalisation –, longer utterances will probably require more complex processing, i.e. they will have to rely more heavily on the combinability constraints of Level 3.[14]

The 'interwoven' fragments of (26) provide another example of how discontinuously combined utterances can yield fairly complex structures, a crossing-dependency construction in the example. The corresponding coverage structure is given as Table 12.

(26)

INTERWOVEN FRAGMENTS (Crossing dependency)
Fragment1: Anna ran Fragment2: Betty jogged
Fragment3: Charles walked Fragment4: Betty and Charles
Fragment5: jogged and walked respectively

Anna, Betty, and Charles ran, jogged, and walked, respectively

Table 12

Anna	*Betty*	*and*	*Charles*	*ran*	*jogged*	*and*	*walked*	*respectively*
Anna				ran				
	Betty				jogged			
			Charles				walked	
	Betty	and	Charles					
					jogged	and	walked	respectively

Recall that our coverage experiments were meant to demonstrate that fragment combinations may span relatively complex utterances. We emphasise, however, that no commitment was made as to the grammatical 'correctness' of any of the fragments. As the example of (12), repeated as (27) a), suggests, there can be several ways to achieve the same degree of

[14] Cf. our similar arguments in Section 6.3 concerning the results of Dąbrowska (2008) on English long-distance dependency constructions.

coverage. It remains 5/6 = 83% for 27) b) and c) where fragments *we do* and *do some* are missing, respectively. A learner will most probably develop his\her system of allowable combinations in such a way that processing be optimal – in some sense.

(27)

a) shall we do some drawing then

b) shall we do some drawing then

c) shall we do some drawing then

4.2. *Language acquisition and generalisation*

An interesting phenomenon in children's language acquisition is overgeneralisation. When children seem to have acquired some "rules" for "generating" regular well-formed utterances, they also tend to apply those rules where they should not. Brown (1973, p. 333), for instance, reports mistakes like in (28).

(28)

a) *He goed to make another one.
b) *Did you bought this?
c) *What did he doed?

In (28) a) the regular *–ed* past form is incorrectly generated for *go*, in b) the past tense form *bought* is correct but it is incorrectly used for *buy*, in c) both the application of past tense and the actual past tense form are wrong. Let us examine how such mistakes can follow from an agreement groups model. Suppose the child has stored the *you play* group of (29) in memory.
The group contains both present and past forms of both regular and irregular verbs. Note that all the novel utterances, *we help, we fall, we helped, we fell*, compatible with the group are grammatical.

(29)

you play	we play	you help	you fall
you played	we played	you helped	you fell

It is reasonable to expect the child to realise, at some point of development, that the words within the *{play, help, fall, played, helped, fell}* category can be differentiated with respect to their endings. Some words have *–ed*, some do not. This may entail, on the one hand, the division of *{play, help, fall, played, helped, fell}* into categories *{play, help, fall, fell}* and *{played, helped}*. On the other hand, since *–ed* can now be seen as a speech fragment that can combine with the "stems" of the *{played, helped}* category words, it may be necessary to revise the grouping of (29) because originally it consisted of two-unit utterances but now some utterances seem to consist of three units. Thus the two groups of (30) seem to be a more rational distribution of the original utterances.

(30)

<u>you play</u>	we play	you help	you fall	you fell
<u>you play ed</u>	we play ed	you help ed		

Next, the learner can notice that the *{play, help}* category of the *Y X ed* frame is actually a subset of the *{play, help, fall, fell}* category of the *you play* group and conclude that *{play, help, fall, fell}* category words are allowable for the *you play ed* group, too. Naturally, this would lead to such overgeneralisation errors as shown in (31). The *falled* verb form parallels *goed* in (28) a) and *doed* in c). The *felled* form predicted by our approach is even more interesting since it may correspond to an intermediate developmental state where the analytical *–ed* past form "fights" the item based irregular one, and the outcome is a draw. Such verb forms do exist in children's language: Brown (1973, p. 335), for instance, explicitly refers to *tored, broked* and *felled.* Note that such kinds of errors are linked primarily to the first level of generalisation in our model as they can be accounted for in terms of agreement groups and agreement categories. *Did you bought this?* can be explained similarly. Utterances *you buy* and *you bought* may belong to the same group, so they can share an agreement category, say *X={...buy, bought, ...}*. On the basis of some inter-group parallelism the child may erroneously infer that any *X* category element can be used for another category in another frame, say for *Y* in *Did you Y this?* This frame would license both *Did you buy this?* and **Did you bought this?*

(31)

a) Group: Y X ed Y = *{you, we}*
X = *{play, help, fall, fell}*

b) Novel utterances:
*you fall ed *you fell ed we help ed
*we fall ed *we fell ed

The ungrammaticality of **What did he doed?* probably involves some interference from the covering mechanisms of the higher generalisation levels. Suppose the child has two agreement groups corresponding to frame *Y X ed* and frame *What did Y X*, respectively, with *Y* including *he* and *X* including *do*. Figure 11 shows how the two groups can cover the erroneous utterance. The resultant coverage structure in Figure 11 can constitute a candidate schema or combinability constraint as illustrated in Table 13. Whether or not the coverage structure develops into a fully fledged schema, or CC, might be determined by how optimal it is with respect to its linguistic-communicative function.

Group 1: Y X ed Y = *{you, we, he}*
X = *{play, help, fall, fell, do}*

Group 2: What did Y X

Coverage structure:

What	*did*	*he*	*do*	*ed*
What	did	he	do	
		he	do	ed

Figure 11. Hypothetical coverage structure for **What did he doed?* effected by Group1 and Group2

Table 13

What	did	he	do	ed
2_1	2_2	2_3	2_4	
		1_1	1_2	1_3

The above examples also indicate that the agreement groups method is not confined to a single linguistic module. Our analysis involved the past tense morphology of English verbs, thereby traversing the syntax-morphology interface of language.
The fact that overgeneralisation errors minimize as the child becomes older relates to some underlying error correction mechanisms. In our framework this broadly means the regrouping of utterances (i.e. homogenising groups), reorganising schemas, or revising combinability constraints for fragments. Viewed from a statistical perspective, such corrections can be connected to the notions of *preemption* – "repeatedly witnessing the word in a competing pattern" – and a pattern's *degree of openness* – "the variability of the items that occur in a given pattern" – as advocated by e.g. Goldberg (2009a). Of the many possible alternative groups, schemas, and CCs, those ones will be "finalised" (long-term optimization) or applied in the current speech act (situation-level optimisation) which have been witnessed[15] most often in previous similar communicative situations (preemption) or are compatible with the most linguistic items (degree of openness). Tomasello (2003) considers *entrenchment,* besides preemption, an important constraint on generalisation: "Entrenchment simply refers to the fact that when an organism does something in the same way successfully enough times, that way of doing it becomes habitual and it is very difficult for another way of doing that same thing to enter into the picture" (p. 300). From our point of

[15] i.e. have been used successfully for processing,

view, only those groups, schemas, and CCs should survive or be dominant which have been used successfully in numerous communicative situations.
Tomasello (2003, Section 4.3.4) characterise children's early syntactic development in terms of the construction types that are available at subsequent developmental stages. The constructions for each later stage become more complex due to the appearance of more advanced cognitive functions. First, holophrases do not exhibit any relevant syntactic features. Second, word combinations ("pivot schemas") can partition events lexically. Third, item-based constructions can mark participant roles syntactically, but syntactic performance depends strongly on the particular constructions. Finally, abstract constructions appear that characterise events in terms of participant roles in a more generic fashion. Such abstract "schemas" can be represented by category sequences like e.g. *trans-SUBJ trans-VERB trans-OBJ* meaning that the subject of a transitive verb precedes the verb in question and the object comes after the verb. Roughly, we can say that the developmental process consists in going from holophrases like *mummy-loves-you* to category sequences like *SUBJECT VERB OBJECT*, or more generically, *NOUN VERB NOUN*. Let us see how this can be done in the agreement groups framework.
First the child has to discover that holophrases are made up of words. The discovery process may be related to what we said about agreement groups construed as superpositions of frames in Section 2.4. Given input utterances (32) a) the child may "notice" that they can be grouped so that the utterances in a group have a fixed part and a variable part. Cf. (32) b)-d). The fixed parts can serve as the foundations of item-based phrases, whereas variable parts form classes of linguistic units that can fit into the slots of such frames. These variable linguistic units will be the words in the child's vocabulary. Thus segmentation of holphrastic utterances and group formation may be parallel processes.

(32)

a)	mummy-loves-you	daddy-loves-you	Adam-loves-you
	mummy-needs-you	mummy-kisses-you	mummy-loves-him
	mummy-loves-daddy		

b)	**mummy**-loves-you	item-based phrase: X-loves-you
	daddy-loves-you	X={mummy, daddy, Adam}
	Adam-loves-you	
c)	mummy-**loves-**you	item-based phrase: mummy-Y-you
	mummy-**needs-**you	Y={loves, needs, kisses}
	mummy-**kisses-**you	
d)	mummy-loves-**you**	item-based phrase: mummy-loves-Z
	mummy-loves-**daddy**	Z={you, daddy, him}
	mummy-loves-**him**	

As a next step, the child "realises" that even the fixed parts of item-based phrases are composed of units that function as variable parts for some item-based phrases, i.e. that utterances are actually word sequences. At this point the child can start to look at the utterances coming from the environment in a different way: he/she is ready to look for word sequences in the speech stream. At the same time the child can start to build syntactic representations. This can be done in two, basically equivalent, ways, both ways involving the restructuring of the inventory of stored linguistic material. First, the child can reinterpret stored chunks as actually composed of smaller units, i.e. words belonging to word classes established on the basis of item-based phrases. This can entail the regrouping of the original utterances in terms of some similarity criterion, say our one-word difference requirement. For example, the child could recognise that utterances (32) a) meet the one-word difference criterion with respect to base sentence *mummy loves you*, and “treat” them as a group. The tabular representation of the group, in turn, could serve Level 1 generalisation purposes.

An equivalent way of deriving a tabular representation is to regard item-based phrases as frames and superimpose them (cf. Section 2.4). The superposition of the item-based phrases of (32), for instance, would yield category sequence *X Y Z* which in turn would represent a hypothetical table for the agreement group corresponding to (32) a) with agreement categories *X*, *Y*, and *Z* standing for the respective columns in the table. What is left to reach Tomasello’s (2003) abstract constructions stage is to identify *X* with subject (or noun), *Y* with verb, and *Z* with object (or noun). A precise answer to the question of how this is exactly done may lie outside the scope of our model. However, we note that there can be immediate ways of assigning syntactic categories to agreement categories. Consider the group in (33). Since *what* is in the same category as *it, penguin,* and *that*, it can explicitly indicate for the child that *it, penguin,* and *that* are “nominal” words, i.e. *Z* = “noun” for frame *W X Y Z*. Furthermore, if *do* was regarded by the child as a word prototypically denoting some action, it could be inferred that *cuddle* and *eat* also stand for some action, i.e. *Y* = “verb” for frame *W X Y Z*.[16]

(33)

you can do it	you can do that	you can cuddle it	you can't do it
you can do penguin	Anne can do it	you can eat it	you can do ***what***
I can do it			

This direct category assignment may disagree with what Langacker (2009, p. 173) says in drawing parallels between Cognitive Grammar (CG, Langacker 1987), Cognitive Construction Grammar (CCxG, Goldberg 2006), and Radical Construction Grammar (RCxG, Croft 2001). He claims that “… CG, CCxG, and RCxG agree that distributional classes do not provide the basis for general characterizations of notions like noun, verb, subject, and object. Even in a single language, there may be no construction in which appear all and only those elements commonly recognized as nouns or verbs”. The example of (33) illustrates that it may not be absolutely necessary for, say, all nouns to be compatible with “construction” (33) in order that their

[16] Some further examples can be found in Drienkó (2014)

agreement category or "distributional class" be allotted "metacategory" *WHAT*, i.e. ca. *NOUN*.

Waterfall et al. (2010) and Waterfall & Edelman (2009) emphasise the relevance of "variation sets" in child-directed speech as discourse cues for language acquisition. A variation set is "a contiguous sequence of utterances produced by a single speaker in a conversation and each successive pair of utterances has a lexical overlap of at least one element" (Waterfall et al. 2010, p. 687). It is easy to see some kinship between agreement groups and specific variation sets: for instance, the possible sequence of utterances *this is my doll, this is your doll, there is my doll, this is my doggie* constitutes a variation set but the utterances can form an agreement group, as well, since each utterance differs from *this is my doll* in only one word. Variation sets, thus, might facilitate group formation, the fundamental component of our generalisation model.

4.3. Generalisations at work

We singled out three basic cognitive levels for linguistic generalisations and associated with each level a characteristic type of generalisation object: agreement groups with Level 1, schemas with Level 2, and combinability constraints with Level 3. Arguably, there may also be a Level 0 of holophrases which can be characterised with the lack of syntactic generalisations. The computational processes associated with each level were said to build on the abstract representations of previous levels. So far, we have mostly referred to three types of computational processes: i) combination of words ("holophrases") within agreement groups, i.e. processing novel utterances compatible with a given agreement group, ii) combination of agreement groups (fragments) within schemas, and iii) combinations of agreement groups (fragments) transcending the limits of schemas (CCs). In principle, however, generalisation should not be confined to just these processing types: it may be possible to combine each generalisation object with any other(s).

Thus, in general, we might have computational processes involving holophrase-holophrase, holophrase-group, holophrase-schema, group-group, group-schema, and schema-schema combinations. It would also be plausible to assume that in processing a given utterance several of these combination mechanisms may be involved simultaneously.

To see a holophrase-holophrase combination, for instance, consider such holophrastic expressions as *I-think* and *it's-OK*. The utterance *I think it's OK* could then be interpreted as the juxtaposition of the two expressions.[17]

Dąbrowska (2008) analysing English long-distance dependency (LDD) constructions from a usage-based perspective reports that the most acceptable interrogative LDD utterances were those which fit either template *WH do you think S-GAP?* or *WH did you say S-GAP?*. Furthermore,

[17] Note that our model does not provide an immediate principle for distinguishing between multi-word utterances and holophrases. However, it would be quite straightforward to regard utterances above a certain frequency threshold as holophrases and assign some 'holophrase' category to them.

i) LDD questions disprefer other main verbs than *think* or *say*, other auxiliaries than *do*, as well as complementisers; ii) LDD questions virtually never involve dependencies longer than one clause. The data can be explained with reference to differences in processing levels. Suppose we have agreement groups (34) a) and b) and holphrases (34) c). The utterances of (35) a) can be obtained by the combination of group (34) a) and holophrastic expression *do-you-think*, whereas combining (34) a) with holophrase *did-you-say* yields sentences (35) b). Note that combination here may simply mean ”insert”. Combining group (34) a) and *I-think* leads to the “add on” utterances of (35) c) and d).

(34)

a)	b)	c)
When Anne eats it	do you think	do-you-think
How Anne eats it	can you think	did-you-say
When she eats it	do you say	I-think
When Anne drinks it	do they think	
When Anne ate it	did you think	

(35)

a)
When do-you-think **Anne eats it**
When do-you-think **Anne ate it**
How do-you-think **she drinks it**

b)
When did-you-say **Anne eats it**
When did-you-say **Anne ate it**

c)
I-think **When Anne eats it**
I-think **When she ate it**

d)
When Anne eats it I-think
When she ate it I-think

However, for adult speakers, *do you think* or *did you say* can also be available as word sequences that are members of corresponding agreement groups. Cf. (34) b). The fragment-fragment (i.e. group-group) type combination of (34) a) and b) results in sentences like (36).

(36)

When can they think **Anne eats it**
When did they think **Anne eats it**
When can they say **Anne eats it**

Since an “insert”, or “add on” type holophrase-group combination may require less computational efforts than the combination of the words from two groups, *When do-you-think Anne eats it*, for example, may be processed more readily – i.e. is “more acceptable” – than *When can they think Anne eats it*. The holophrastic interpretation of *do-you-think*, or *did-you-say* also explains why no other words are welcome in the LDD constructions. Why LDD constructions do not prefer longer dependencies may be explained by the fact that the involvement of even more fragments would require even more computation. Alternatively, we could also think of utterances like (35) a) and b) as representing the underlying *WH do you think X Y it* or *WH did you say X Y it* schemas, respectively. Further clauses, then, would require

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

stepping out of the schemas, which would entail the need for the more complex computational processes of Level 3.
Allowing the combination of any type of generalisation object with any other entails that the combinability constraints of Level 3 should possibly involve not only fragment-fragment (group-group) combinations but also holophrase-holophrase, holophrase-group, holophrase-schema, group-group, group-schema, and schema-schema combinations. The extension of combinability constraints for any objects, in turn, requires the reformulation of the mechanisms of extracting CCs from generalisation objects. Instead of "schema", we will use the more general term "matrix object", i.e. an object containing object(s). Recall we have the following types of generalisation objects: holophrase (Level 0), group (Level 1), schema (Level 2), and constraint (Level 3). A holphrase is considered to be a single unit, though it may contain several words. An agreement group consists of utterances minimally contrasting with a base utterance. A schema is an abstract coverage structure showing how certain more complex utterance types can be obtained by combining certain groups. A constraint is a partial coverage structure showing how two objects can be combined in general. Then, a combinability constraint $CC(obj_1, obj_2)_k$ for generalisation objects O_1 and O_2 in matrix object *MO* is a sequence of agreement categories $c_1 \ldots c_n$, where each c_i ($1 \leq i \leq n$) occupies a unique position in *MO*, each c_i belongs to either object O_1 or O_2, and if c_i precedes c_j in *MO* then c_i precedes c_j in $CC(obj_1, obj_2)_k$. If c_i and c_j appear in the same column of *MO* their order is immaterial in $CC(obj_1, obj_2)_k$.[18] Agreement categories appearing in different columns of *MO* are separated by commas.
In our model, language acquisition can be seen as a process of continuously extracting patterns, i.e. generalisation objects, from utterances coming from around us. The extraction process (acquisition), in principle, could go on for the entire lifetime of the learner. This is reflected in the fact that our definition of CCs allows CCs operating with CCs. That is to say, combinability constraints – i.e. partial coverage structures – can be combined to obtain further combinability constraints – larger (partial) coverage structures. Such "higher level" constraints may be useful for processing frequently encountered speech fragments. They can speed up processing since they can function as prefabricated building blocks for still higher-level processing due to the fact that the objects which they consist of need not be combined from scratch. On the other hand more and more constraints requiring more and more memory space would eventually overburden the cognitive system. This suggests a kind of trade-off between schematisation – i.e. creating and storing abstract generalisation objects – and the "online" cognitive mechanisms subserving linguistic processing – i.e. mapping on groups, and coverage. Accordingly, storing both matrix objects and the combinability constraints extracted from them may be useful initially, but later on, having an abundant inventory of stored linguistic examples, it may be more economical to "remember" only those novel generalisation object combinations which yield significant benefit, e.g. by

[18] For simplicity the agreement category represented by a smaller number is written first.

speeding up processing. For instance, it may be unnecessary to keep in memory a novel coverage structure when it can be obtained by simply combining an already existing schema and a group. It would be more economical to record the novel structure as a novel combinability constraint for the two objects in question.

In (37) we reformulate the coverage parsing algorithm of (21) in terms of the more general notion "generalisation object" instead of "fragment".

(37)

1. Given: input utterance U
2. Cover the utterance with possible generalisation objects, i.e. set up coverage structure CS for U
3. Find a coverage structure CS_k (subset of CS) consisting of k objects arranged in k rows in CS such that any two objects, O_i and O_j ($1 \leq i, j \leq k$, $i \neq j$) in CS_k be combinable, i.e. for any pair of objects there exist a combinability constraint $CC(i,j)$ or $CC(j,i)$.
4. U is grammatical if CS_k exists (i.e. if every word position in U is covered by CS_k)

Again, we can relax the combinability requirement in the 3rd step of the algorithm to effect the processability of utterances that cannot be covered so "densely". For example, given k objects, it might be sufficient to require k-1 distinct combinability constraints instead of the $k(k-1)/2$ ones required by step 3 of (37).

To illustrate the interplay of generalisation objects in linguistic processing, let us demonstrate how sentence *I think Eve Anna Betty and Charles cycled ran jogged and walked respectively* might be covered on the basis of the schema in Figure 12 – representing the basic 'respectively' construction – and three combinability constraints derivable from three respective matrix objects.

Betty	and	Charles	jogged	and	walked	respectively
Betty			jogged			
		Charles			walked	
Betty	and	Charles				
			jogged	and	walked	respectively

Schema1: 'respectively' construction

Figure 12. Coverage structure representing a schema for the 'respectively' construction

The coverage structure in Figure (13) for *I think Betty and Charles jogged and walked respectively* is a matrix object from which combinability constraint *CC1* can be extracted licensing the combination of holophrase *I think* and the schema underlying utterance *Betty and Charles jogged and walked respectively.*[19]

[19] Recall that coverage structures consist of categories representing groups, but we write the actual words of the utterances in question for readability. The discontinuous lines in the coverage structure indicate that *I think* as a holophrase should be regarded as a single item and be represented by a single column.

I	think	Betty	and	Charles	jogged	and	walked	respectively
		Betty			jogged			
				Charles			walked	
		Betty	and	Charles				
					jogged	and	walked	respectively
I	think							

Holophrase1: I-think
Schema1: 'respectively' construction → CC1(Holophrase1, Schema1)

Figure 13. Matrix object for extracting combinability constraint CC1.

The coverage structure for *Anna Betty and Charles ran jogged and walked respectively* provides *CC2* which enables the combination of the group behind *Anna ran* and the schema symbolised by Figure 12. Cf. Figure 14. Finally, *Eve and Anna cycled and ran*, the matrix utterance of Figure 15 supplies *CC3* for combining the two hypothetical groups underlying *Anna ran* and *Eve cycled*, respectively.

Anna	Betty	and	Charles	ran	jogged	and	walked	respectively
Anna				ran				
	Betty				jogged			
			Charles				walked	
	Betty	and	Charles					
					jogged	and	walked	respectively

Group1: 'Anna ran'
Schema1: 'respectively' construction → CC2(Group1, Schema1)

Figure 14. Matrix object for extracting combinability constraint CC2

Eve	and	Anna	cycled	and	ran
		Anna			ran
Eve			cycled		
Eve	and	Anna			
			cycled	and	ran

Group1: 'Anna ran'
Group2: 'Eve cycled' → CC3 (Group1, Group2)

Figure 15. Matrix object for extracting combinability constraint CC3

Our coverage parsing algorithm can now establish coverage structure Table 14 for input utterance *I think Eve Anna Betty and Charles cycled ran jogged and walked respectively*.

Table 14

I	*think*	*Eve*	*Anna*	*Betty*	*and*	*Charles*	*cycled*	*ran*	*jogged*	*and*	*walked*	*respectively*
			Anna					ran				
				Betty					jogged			
						Charles					walked	
				Betty	and	Charles						
									jogged	and	walked	respectively
		Eve					cycled					
I	think											

The next step is to find combinable objects within the coverage structure. *CC1* permits the combination of *I think* and *Betty and Charles jogged and walked respectively*. Cf. Table 15.

Table 15

I	*think*	*Eve*	*Anna*	*Betty*	*and*	*Charles*	*cycled*	*ran*	*jogged*	*and*	*walked*	*respectively*
			Anna					ran				
				Betty					jogged			
						Charles					walked	
				Betty	and	Charles						
									jogged	and	walked	respectively
		Eve					cycled					
I	think											

Table 16 shows how *CC2* allows the combination of *Anna ran* and *Betty and Charles jogged and walked respectively*.

Table 16

I	*think*	*Eve*	*Anna*	*Betty*	*and*	*Charles*	*cycled*	*ran*	*jogged*	*and*	*walked*	*respectively*
			Anna					ran				
				Betty					jogged			
						Charles					walked	
				Betty	and	Charles						
									jogged	and	walked	respectively
		Eve					cycled					
I	think											

The combination of *Eve cycled* and *Anna ran* is possible via *CC3* as illustrated in Table 17. The grey cells indicate that every position of the utterance is covered by some object.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

Table 17

I	*think*	*Eve*	*Anna*	*Betty*	*and*	*Charles*	*cycled*	*ran*	*jogged*	*and*	*walked*	*respectively*
			Anna					ran				
				Betty					jogged			
						Charles					walked	
				Betty	and	Charles						
									jogged	and	walked	respectively
		Eve					cycled					
I	think											

The coverage structure for our example utterance involves four generalisation objects: i) the schema underlying *Betty and Charles jogged and walked respectively*, ii) Group 1 licensing *Anna ran*, iii) Group 2 licensing *Eve cycled*, iv) holophrase *I think;* and three combinability constraints, *CC1, CC2,* and *CC3*. Thus the utterance can be considered grammatical if our algorithm adopts the less strict parsing condition, i.e. *k-1* constraints for *k* objects. Note that our example schema was treated as a single unit, i.e. as a single row in the coverage structures, as far as combinability was concerned. This may reduce computational costs significantly, since combinability is not checked for "sub-objects" of objects.

4.4. Agreement groups as usage-based constructions

In Fillmore (1988, p. 54) *"grammatical constructions* [are treated] *as syntactic patterns which can fit into each other, impose conditions on each other, and inherit properties from each other. Grammatical constructions define positions...".* This view is largely consonant with our picture of agreement groups construed as representing simple or "primary" constructions which can be combined into more abstract schemas. Due to its usage-based stance, however, explicit a priory defined feature ontologies, feature-specifications, or conditions/constraints are not part of our model. On the other hand, agreement groups may, indeed, symbolise linguistic constructions as (38) and (39) show. Recall also the relationship between utterance positions and our agreement categories. (Cf. Section 2.2)

(38)
'That's X isn't it' construction (Group 2924)

<u>that's right isn't it</u>	that's easier isn't it	that's better isn't it	that's lovely isn't it
that's good isn't it	that's Roger isn't it	that's nice isn't it	that's Joseph isn't it
that's alright isn't it	that's blue isn't it	that's bright isn't it	

(39)
'That's a X Y' construction (Group 10127)

<u>that's a funny one</u>	that's a funny laugh	that's a funny picture	that's a big one
that's a good one	that's a red one	that's a funny man	that's a funny pussy
that's a Daddy one	that's a funny noise	that's a smaller one	that's a hard one
that's a old one	that's a little one	that's a funny thing	that's a funny song
that's a new one	that's a lemon one		

Furthermore Figure 16 demonstrates how agreement properties can be "inherited" in utterances covered by the two constructions. The 3rd–singular agreement between *that's* and *isn't it*, characteristic of the *that's right isn't it* group, is also valid for utterance *that's a good thing isn't it.*

that's good isn't it: 2924_ 1 2924_ 2 2924_ 3 2924_ 4
that's a good thing: 10127_ 1 10127_ 2 10127_ 3 10127_ 4

That's	*a*	*good*	*thing*	*isn't*	*it*
10127_ 1	10127_ 2	10127_ 3	10127_ 4		
2924_ 1		2924_ 2		2924_ 3	2924_ 4

Figure 16. 3rd-singular agreement "inherited" in utterance *that's a good thing isn't it* covered by groups 2924, and 10127.

Discussing Subject-Auxiliary Inversion (SAI), Goldberg (2009b, p. 211) asks the question "whether the set of constructions that exhibit SAI are best simply stipulated, or whether instead they should be recognized to form a network of related cases." The agreement groups counterpart of such a question might be something like: Are SAI constructions represented by individual groups (fragments) or can they be effected by the combination(s) of more elementary fragments? Our model seems to support the former option as long as '*isn't*' is considered to be a single word. Groups a) and b) in (40), for instance, could straightforwardly be regarded as representatives of the respective constructions accounting for the data in c).

(40)

a)	b)
<u>he is faster than she is</u>	<u>he is faster than is she</u>
he is faster than John is	he is faster than is John
he isn't faster than she is	he is slower than is she
he is slower than she is	Anne is faster than is she
Anne is faster than she is	

c)

he is faster than she is	← a)
he isn't faster than she is	← a)
he is faster than is she	← b)
*he isn't faster than is she	neither a) or b)

On the other hand, it is difficult to imagine how the data could be covered by the combination of more elementary fragments. To illustrate, the combinability of the *he X faster* group with the *faster than she is/faster than is she* groups in Figure 17 depends on whether *X* is negated or not. But what can prevent *he isn't faster* from being a member of the *he X faster* group, given only our one-word difference criterion for group formation? If, then, *he*

X faster could be combined with the *faster than is she* group the ungrammatical **he isn't faster than is she* could follow.

Group1: he is faster	Group2: faster than she is	Group3: faster an is she
he isn't faster	faster than he is	faster than is he
he is slower	slower than she is	slower than is she
…	…	…

He	*is*	*faster*	*than*	*she*	*is*
He	is	faster			
		faster	than	she	is

He	*isn't*	*faster*	*than*	*she*	*is*
He	isn't	faster			
		faster	than	she	is

He	*is*	*faster*	*than*	*is*	*she*
He	is	faster			
		faster	than	is	she

**He*	*isn't*	*faster*	*than*	*is*	*she*
He	isn't	faster			
		faster	than	is	she

Figure 17. Combining Group1 with Group2 yields grammatical utterances, whereas Group1 and Group3 produces **he isn't faster than is she.*

By interpreting '*isn't*' as consisting of '*is*' and '*not*', i.e. two words, we can arrive at a different conclusion. Now we have three groups, (41) a)-c), involved in the account for the data under (41) d). However, in this case it is also possible to obtain a coverage account since one may find a negated fragment F_{Not} licensing e.g. *he is not faster*, combinable with the uninverting fragment $F_{she\text{-}is}$ licensing e.g. *faster than she is* but incombinable with the inverting fragment $F_{is\text{-}she}$ responsible for e.g. *faster than is she*. In other words, our coverage parsing algorithm detects a valid coverage structure for *he is not faster than she is* because $CC(F_{Not}, F_{she\text{-}is})$ exists, but due to the non-existence of $CC(F_{Not}, F_{is\text{-}she})$ **he is not faster than is she* will be unacceptable. Cf Figure 18. The positive sentences, *he is faster than she is*, and *he is faster than is she* can be accounted for by a positive fragment F_P and constraints $CC(F_P,F_{she\text{-}is})$ and $CC(F_P,F_{is\text{-}she})$.

(41)

a)
he is faster than she is
he is faster than John is
he is slower than she is
Anne is faster than she is

b)
he is faster than is she
he is faster than is John
he is slower than is she
Anne is faster than is she

c)
he is not faster than she is
he is not faster than John is
he is not slower than she is
Anne is not faster than she is

d)
he is faster than she is ← a)
he is not faster than she is ← c)
he is faster than is she ← b)
*he is not faster than is she neither a) or b) or c)

Group1: (F_{Not}) he is not faster
he is not slower
he was not faster
…

Group2: ($F_{she\text{-}is}$) faster than she is
faster than he is
slower than she is
…

Group3: ($F_{is\text{-}she}$) faster than is she
faster than is he
slower than is she
…

He	*is*	*not*	*faster*	*than*	*she*	*is*
He	is	not	faster			
			faster	than	she	is

**He*	*is*	*not*	*faster*	*than*	*is*	*she*
He	is	not	faster			
			faster	than	is	she

Figure 18. Utterance '*he is not faster than she is*' is grammatical because Group1 and Group2 can be combined via CC(F_{Not}, $F_{she\text{-}is}$), whereas **he is not faster than is she* is unacceptable due to the non-existence of CC(F_{Not}, $F_{is\text{-}she}$), i.e. because Group1 and Group3 cannot be legally combined.

Popularizing Fluid Construction Grammar, van Trijp et al. (2012) underline the importance of the bidirectionality or reversibility of their implementation, i.e that a single linguistic inventory is fit both for parsing and producing utterances. Note that agreement groups can also be used in both ways. Our "mapping on groups", is actually a kind of parsing, especially so when mapping is understood as "coverage" and the output is the "coverage structure" of the input utterance. On the other hand, as we noted in Section 2.2, agreement groups are a powerful device for processing novel sentences. Naturally, processing might mean not just mapping onto groups, but it is also possible to produce utterances by selecting arbitrary elements with appropriate agreement categories and string them together as constrained by the generalisation objects of our model.

Journal of Child Language Acquisition and Development – JCLAD
Vol: 4 Issue: 3 113-158, 2016,September
ISSN: 2148-1997

The Embodied Construction Grammar of Bergen & Chang (2013) views constructions as an interface to mental simulation. We suspect that our framework does not contradict such a view since whatever the actual neurological representations of constructions – basically, groups of utterances, in our case – are, the formation of groups should not prevent group-member utterances (and/or the words they are composed of) from having access to or being accessible by other brain modules, the sensory-motor system, for instance.

Sag et al. (2012) divide their Sign-Based Construction Grammar framework, "*meant to be compatible with most linguistic analyses that have been developed within CxG of all kinds*" (p. 19), into three main parts: signature, lexicon and constructicon. Broadly, our agreement groups model might seem to have such components. Our "lexicon" contains words (holophrases) together with their agreement categories. The "constructicon" in our model consists of generalisation objects. "Signature" for us might mean an inventory of fundamental notions and procedures for the characterisation of the various components and processes underlying the generalisation mechanisms of mapping on groups and schemas, assembling coverage structure, extracting combinability constraints, and coverage "parsing".

5. Conclusions

The results summarised in this work might possibly be understood as some evidence that linguistic analyses based on a suitable grouping of utterances, in particular on agreement groups as defined for our method, may be informative with respect to the cognitive mechanisms underlying linguistic behaviour. We showed that agreement groups can account for both individual utterances and utterances functioning as fragments for longer utterances. It was also demonstrated how the two concomitant cognitive capacities – grouping, and finding smaller fragments in larger ones – can serve as a basis for a continuum model of linguistic generalisation. The model, assuming three major levels of abstraction, suggests the following order of development for generalisation objects: holophrase, group, schema, constraint. The notion of "coverage structure" was extensively used throughout our analyses and it provided a principal footing for our formal statements about schemas, combinability constraints, and "coverage parsing".

Our approach is *distributional* in that the emergence of generalisation objects, especially the grouping of utterances, is directly determined by the distributional properties of the "training set/corpus". Our approach is *usage-based* since the emergence of linguistic objects is critically determined by the actual linguistic material the learner encounters. Our model is *"computationist"* as we consider the computational mechanisms underlying linguistic processing at least as important as the linguistic objects themselves. Our framework is also *constructionist* because it enables explicit correspondences between individual constructions and generalisation objects.

In order to counterbalance grouping errors due to noise, ungrammaticality and ambiguity in the training corpus we assume some error correction mechanisms that can delete mischievous utterances from misbehaving groups, or can delete an incorrect group altogether. Alternatively, the corpus can be preprocessed so that unwanted effects cannot interfere with grouping. Note, however, that our algorithms did not employ any kind of supervision. This may seem to be in line with various computational results showing signs of the possibility that unsupervised statistical learning of grammar representations is possible from non-annotated corpora (Solan et al. 2005, Bannard et al. 2009, Bod 2009, Waterfall et al. 2010). Nevertheless, we suspect that the formation of "perfect", or, in our terminology, "homogeneous" generalisation objects requires some feedback, i.e. supervision, since the inherent ambiguities, and errors of human speech might not be entirely captured by statistical techniques.

References

Babarczy, A. (2006). The development of negation in Hungarian child language. *Lingua* 116, pp. 377-392.

Bannard, C., Matthews, D. (2008). Stored Word Sequences in Language Learning. *Psychological Science,* 19(3), 241-248.

Bannard, C. Lieven, E. Tomasello, M. (2009). Modeling children's early grammatical knowledge. *PNAS Vol 106, no. 41.* 11629-11634.

Bergen, B., Chang, N. (2013). Embodied Construction Grammar. In Thomas Hoffmann and Graeme Trousdale (Eds.) *Oxford Handbook of Construction Grammar,*. Oxford University Press.

Bod, R. (2009). From Exemplar to Grammar: A Probabilistic Analogy-Based Model of Language Learning. *Cognitive Science* 33, 752–793

Brown, R. (1973). *A first language. The early stages.* Harvard University Press, Cambridge, Massachusetts

Cameron-Faulkner, Th., Lieven, E., Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science.* 27. 843-873

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.

Dąbrowska, E., Lieven, E.(2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16–3, 437–474

Dąbrowska, E. (2008) Questions with long-distance dependencies: A usage-based perspective
Cognitive Linguistics, 19(3). DOI:10.1515/COGL.2008.015

Drienkó, L. (2004a). Agreement Mapping System Approach to Language. *Journal of Language and Linguistics.* Vol. 3. No. 1. 38-61. http://webspace.buckingham.ac.uk/kbernhardt/journal/3_1/drienko3_1.html

Drienkó, L. (2004b). Outlines of Agreement Syntax. *Journal of Language and Linguistics.* Vol. 3. No. 2. 154-181. http://webspace.buckingham.ac.uk/kbernhardt/journal/3_2/drienko3_2.htm

Drienkó, L. (2012a). Agreement groups analysis of mother-child discourse. Talk presented at the *4th UK Cognitive Linguistics Conference, King's College London, UK,* 10-12th July 2012.

Drienkó, L. (2012b) *A linguistic agreement mapping-system model: agreement relations for linguistic processing*. LAP-Lambert Academic Publishing

Drienkó, L. (2013a). Distributional cues for language acquisition: a cross-linguistic agreement groups analysis. Poster presentation for the *11th International Symposium of Psycholinguistics, Tenerife, Spain* 20-23 March, 2013

Drienkó, L. (2013b). Agreement groups coverage of mother-child language. Talk presented at the *Child Language Seminar, Manchester, UK*, 23-25 June 2013

Drienkó, L. (2013c). Agreement groups coverage of Hungarian mother-child language. Poster presentation for the *11th International Conference on the Structure of Hungarian*. 29-31 August 2013, Piliscsaba, Hungary

Drienkó, L. (2014). Agreement groups analysis of mother-child discourse. In Rundblad, G., Tytus, A., Knapton, O., and Tang, C. (eds.) *Selected Papers from the 4th UK Cognitive Linguistics Conference*. London: UK Cognitive Linguistics Association. pp. 52-67. http://www.uk-cla.org.uk/files/proceedings/Drienko.pdf

Fillmore, Ch. J. (1988). The Mechanisms of "Construction Grammar". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, pp. 35-55

Finch, S., Chater, N., Redington, M. (1995). Acquiring syntactic information from distributional statistics. In: Levy, JP, Bairaktaris, D, Bullinaria, JA, Cairns, P, (eds.) *Connectionist models of memory and language*. (229 - 242). UCL Press: London.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalizations in Language*. Oxford: Oxford University Press.

Goldberg, A. E. (2009a). The Nature of Generalization in Language. *Cognitive Linguistics* 20 1: 93-127.

Goldberg, A. E. (2009b). Constructions work. *Cognitive Linguistics* 20-1, 201-224

Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago, IL, US: University of Chicago Press. xv 384 pp.

Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. *Psychology of Learning and Motivation*, 7, l-41.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar,* vol. 1, *Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, R.W. (2009) Cognitive (Construction) Grammar. *Cognitive Linguistics* 20-1, pp. 167-176

Lieven, E., Behrens, H., Speares, J., Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language.* 30. pp. 333-370.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2:
The Database. Mahwah, NJ: Lawrence Erlbaum Associates

Mintz, T. H., (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, Volume 90, Issue 1, pp. 91-117. doi:10.1016/S0010-0277(03)00140-9 |

Montes, R. (1987). Secuencias de clarificación en conversaciones con ninos (Morphe 3-4): Universidad Autónoma de Puebla.

Montes, R. G. (1992). *Achieving understanding: Repair mechanisms in mother–child conversations.* Unpublished doctoral dissertation, Georgetown University.

Pinker, S. (1979). Formal models of language learning *Cognition,* 7, 217-283

Redington, M.,Chater, N.Finch, S.(1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. *Cognitive Science* Vol. 22 (4) pp. 425-469

Réger, Z. (1986). The functions of imitation in child language. *Applied Psycholinguistics* 7.323–352.

Sag, I. A., Boas, H. C., Kay, P. (2012) Introducing Sign-Based Construction Grammar
In H. C. Boas & Ivan A. Sag (eds.): Sign-Based Construction Grammar. CSLI Publications.

Solan, Z, Horn, D, Ruppin, E., Edelman, S. (2005) Unsupervised learning of natural languages. *PNAS Vol 102, no. 33.* 11629-11634.

St. Clair, M. C., Monaghan, P., Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*. Volume 116, Issue 3, pp. 341-360

Stoll, S., Abbot-Smith, K., Lieven, E., (2009). Lexically Restricted Utterances in Russian, German, and English Child-Directed Speech. *Cognitive Science* 33, 75–103

Theakston AL, Lieven EV, Pine JM, Rowland CF. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *J. Child Lang*. 28(1):127-52.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

van Trijp, R., Steels, L., Beuls, K., Wellens, P. (2012). Fluid construction grammar: the new kid on the block. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics* (EACL '12). Association for Computational Linguistics, Stroudsburg, PA, USA, 63-68.

Wang, H., Mintz, T. H., (2010). From Linear Sequences to Abstract Structures: Distributional Information in Infant-direct Speech. In Jane Chandlee, Katie Franich, Kate Iserman, Lauren Keil (eds.) *Proceedings Supplement of the 34th Boston University Conference on Language Development*

Waterfall, H.R., Sandbank, B., Onnis, L., Edelman, S. (2010). An empirical generative framework for computational modeling of language acquisition. *J. Child Lang. 37* 671–703.

Waterfall, H.R., Edelman, S. (2009) The neglected universals: Learnability constraints and discourse cues. *Behav. Brain. Sci. 32.* 471-472.

Weisleder A, Waxman S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *J. Child Lang.* Nov; 37(5):1089-108. Epub. 2009 Aug 24.