LingBaW

A peer-reviewed online journal

czasopisma.kul.pl/LingBaW

LingBaW

Linguistics Beyond and Within

VOLUME 4 (2018)

THE JOHN PAUL II CATHOLIC UNIVERSITY OF LUBLIN KUL 1918

LingBaW

A peer-reviewed online journal *https://czasopisma.kul.pl/LingBaW*

LingBaW

Linguistics Beyond and Within

VOLUME 4 (2018)

THE JOHN PAUL II CATHOLIC UNIVERSITY OF LUBLIN KUL 1918

Editorial Team

Editorial Board

TYPESETTING: Karolina Drabikowska

COVER DESIGN: Wojciech Malec

The journal is peer-reviewed. A list of reviewers is published annually on the journal's website at: https://czasopisma.kul.pl/LingBaW/reviewers.

e-ISSN: 2450-5188

PUBLISHER:
The John Paul II Catholic University of Lublin
Al. Racławickie 14
20-950 Lublin, Poland

PUBLISHING HOUSE:
KUL University Press
ul. Konstantynów 1H
20-708 Lublin, Poland
phone: (+48) 81 454-56-78
e-mail: wydawnictwo@kul.pl

EDITORIAL OFFICE ADDRESS:
Institute of Linguistics
The John Paul II Catholic University of Lublin
Al. Racławickie 14
20-950 Lublin, Poland
e-mail: lingbaw@gmail.com

Articles

LingBaW

Linguistics Beyond And Within 4 (2018), 7-21

Acoustic analysis of monophthongs, diphthongs, and triphthongs in Mandarin for 3- to 5-year-old children with articulatory phonological disorders

Man-ni Chu and Jia-ling Syu
Fu Jen Catholic University, Taiwan

Abstract

Ten 3- to 5-year old children (5M, 5F) who were diagnosed as children with articulatory phonological disorders (CWAPD) and attending a therapy program were recruited to participate in a 'repeat-after-her' experiment. They were asked to produce a total of 85 real Mandarin words, including 28 monophthongs, 41 diphthongs, and 16 triphthongs. The results indicated that CWAPD have no problem producing monophthongs. However, attempts to articulate diphthongs and triphthongs induced more errors. CWAPD showed more errors when producing words with 1st sonorant diphthongs than words with 2nd sonorant diphthongs—this is because the least sonorant segment in the last position is prone to distortion. Similar phenomena were found in other triphthongs, except with /iai/ and /iou/, which did not see deviant pronunciation. Comparing our study to the information provided by two therapists showed that the participating CWAPD encountered difficulties in producing multi-vowel syllables, where the position and sonorant matters. In addition, our results also reveal a similar vowel acquisition order among CWAPD as among normal children.

Keywords: articulatory phonological disorders, monophthong, diphthong, triphthong, Mandarin

1. Introduction

Learning language is an important developmental phase for human beings. All children need to experience and master the phonological development process in order to acquire their first language (L1). Children who do not succeed in acquiring L1 usually need medical intervention from speech therapists. In addition, children who cannot communicate linguistically might also display some mental or physical issues. Conditions such as cerebral palsy or various kinds of intellectual disability may reduce a child's learning ability and may cover situations beyond issues relating to problems with the articulatory organs, including: the size and shape of the oral cavity; the length of the tongue and other issues with the frenulum; the arrangement of the teeth; the bite of the upper and lower jaws; cleft lip and palate; and poor coordination of lips and

tongue. The articulatory phonological disorders in this research rule out nerve damage and cognitive mental and physiological structures. Only those who showed symptoms of articulatory phonological disorders were assessed and it was suggested they receive further language therapy. As such, participation in this research was limited to children with functional articulation disorders as defined by Bernthal, Bankson, and Flipsen (2009). Children with impaired hearing, cerebral palsy, cleft lip and palate, and intellectual disability were not a part of this study.

This study is organized as follows: relevant studies on CWAPD are presented in Section 2; methodology and experimental design illustrating data collection, procedures, and analysis are presented in Section 3; the findings are presented in Section 4; and Section 5 offers interpretation of the results. Conclusions and suggestions for further research are presented in the final section.

2. Previous studies

Problems of speech processing involve pronunciation, place, speed, intensity, and coordination. Articulatory disorders can be defined as those disorders involving difficulty in controlling the action of speaking and voicing some sounds. Van Riper (1978) has pointed out that the most important cause of such issues is that the speaker has poor identification of the sound. The speaker cannot distinguish the difference between correct and erroneous pronunciation through hearing alone. According to the American Speech-Language-Hearing Association (ASHA, 1993) the most frequent types of articulatory disorders include substitution, omission, distortion, and addition. 'Substitution' may be described as a way by which speakers replace one sound with another; 'omission' involves the deletion of sounds in words and sentences; in 'addition' one or more extra sounds are added or inserted into a word. The most complex phenomenon is 'distortion,' where the sound is produced partially correctly, meaning some feature of it is distorted. Here we offer some examples from Mandarin Chinese: [pʰiŋ-tuo] for [pʰiŋ-kuo] (**蘋果** 'apple'), where [t] replaces [k] (substitution); [ue] for [uei] (**喂** 'hello'), where [i] is not produced (omission); and [lia-tsɨ] for [ia-tsɨ] (**鴨子** 'duck'), where [l] is added (addition). An example of distortion is [ɕan] for [san] (**山** 'mountain'), where [s] is palatalized, but only this feature is different from [ɕ].

A number of studies (on English (Dodd et al., 1989); Cantonese (So and Dodd, 1994); Spanish (Goldstein, 1996); Turkish (Topbas, 1997); German (Fox, 1997); and Putonghua (Mandarin) (Zhu and Dodd, 2000b)) have investigated subgroups of speech disorders, such as articulation, delay, and (in) consistent disorders. All studies revealed similar developmental processes among children with speech disorders and normal children from similar language backgrounds. For example, Zhu et al. (2000b) analyzed 33 Putonghua-speaking children with speech disorders and compared their data across language backgrounds. One of the conclusions they drew was as to the saliency of the components in the language system in determining the order of acquisition. A similar acquisition process was found among normal Putonghua-speaking children. However, none of the above studies has solely explored the behaviors of children with articulation phonological disorders (CWAPD).

Children come to understand and communicate with language during early childhood and the development of their language skills follows a particular process. This process also provides a basis for growth in a child's cognitive learning, human interactions, emotional development, and social adaptation. The articulation of vowels is usually considered to be easiest and is acquired earlier than that of consonants. Concerning Mandarin vowels, Zheng-Fen Zhang and Yu-Mei Zhong (1986) demonstrated that a child's learning of phonemes begins with simple vowels. The order of vowel development starts with monophthongs, which are then followed by diphthongs and the vowel system may be sufficiently mature by the age of 3. In terms of consonantal development, most oral and nasal stops are acquired before the age of 3; laterals and most fricatives are acquired by age 4; and affricates are acquired after the age of 4. To sum up, the capacity to sufficiently articulate all vowels and consonants has developed by the age of 7. Children's consonantal systems, involving the lips, teeth, and tongue, develop after their vowel systems. Since vowel acquisition is complete by the age of 3, and stop consonants are acquired first, we chose stop consonants to construct a meaningful syllable structure for the study of articulation among CWAPD to avoid the unnecessary influence of vowels. Many studies (Wang et al., 1984; Zhang et al., 1986; Zhu et al., 2000a) have revealed that simple consonants, such as stops, both aspirated and un-aspirated, are acquired before the age of 3. Furthermore, Zhu et al. (2000a) found that triphthongs and diphthongs induced more systematic errors.

According to previous research, articulatory phonological disorders present between the ages of 3 and 5 years old. As such, children who are suspected of having articulatory phonological disorders at age 3 can be assessed by speech therapists. However, coordinated therapy does not usually start until age 4 when children become more cooperative.

Lin (2007) has reported that Mandarin, analyzed as a Consonant-Glide-Vowel-Ending (CGVE) syllable structure, contains: 5 vowels, /i/, /u/, /y/, /a/, and /ə/, where [o] and [ɤ] are allophones of [ə]; 8 diphthongs, [ai], [au], [ei], [ou], [ia], [ie], [ua], [uo]; 5 triphthongs, [iai], [iau], [iou], [uai], and [uei]; and 19 consonants, including /p/, /pʰ/, /t/, /tʰ/, /k/, /kʰ/. In terms of Mandarin vowels, Xu (1987) points out that monophthongs /a/, /o/, /ɤ/, and /e/ develop first, followed by diphthongs /ai/, /ei/, /au/, /ou/, and lastly nasals /an/, /ən/, /aŋ/, /əŋ/. Vowel development is complete before the age of 3. Lin and Lin (1993) suggest that 3-year-old children can pronounce 15 vowels: /a/, /o/, /ɤ/, /e/, /ai/, /ei/, /au/, /ou/, /an/, /ən/, /aŋ/, /əŋ/, /ɚ/, /i/, and /u/, except for /y/.

It is well-known that the first formant, F1, corresponds to vowel openness. F1 is inversely proportional to mouth openness. Open vowels have high F1 frequencies, while closed vowels have low F1 frequencies. The second formant, F2, corresponds to vowel frontness. Back vowels have low F2 frequencies, while front vowels have high F2 frequencies (Pickett, 1999). Since the midpoint of F1 and the F2 vowel segment is a steady area for vowel formants, unaffected by tone, the length of the form-word, and speech tempo (Jeng, 2005), it is reasonable to measure the values of F1 and F2 in order to examine whether the target (openness or the backness of the vowel) has been achieved. Furthermore, Cao (2007) reported a perceptual experiment where the stimuli, 13 diphthongs and triphthongs, were read by 2 males and 2 females from Beijing. She found that the 6:4 duration ratio in the 1st sonorous diphthongs,

where the first vowel is more sonorous than the second vowel, was most recognizable. The ratio in the 2nd sonorous diphthongs was 4:6, and that in the triphthongs was 4:4:2.

The primary goal of this study was to see if acoustic evidence drawn from studying CWAPD, especially on the characteristics of the vowels, revealed a pattern. It would be of interest to further examine CWAPD between the ages of 3 and 5 who speak Mandarin as their first language in terms of both vowels and relatively easy consonants. It was observed that CWAPD needed to generalize the way they produced a particular sound during the therapy process to other settings. This means they need to remember how they produce a particular sound through practice and mimicry. Because CWAPD are at the stage of generalization, their pronunciation is unstable and this may explain differences in their performance between the experiment and therapy. As such, it is reasonable to predict that the least salient element within a syllable may see the largest errors among CWAPD.

3. Methodology

3.1. Stimuli

6 vowels /i/, /u/, /y/, /a/, /o/, and /ɤ/, combined with simple stops /p/, / pʰ /, /t/, / tʰ/, /k/, and /kʰ/, for a total of 28 real monophthongs. 9 diphthongs with the same 6 stops, for a total of 41 real diphthongs, and 5 triphthongs with stops, for a total of 16 triphthongs, as shown in Appendix A.

3.2. Participants and Procedure

14 participations were recruited for the experiment, including: two speech therapists; two normal children (1M (age 5), 1F (age 4.5)); and 10 CWAPD. The CWAPD (5M, 5F, age 3–5) were accompanied by a parent and were asked to repeat after the researcher (the second author) in a quiet room. The stimuli read by the children were recorded using the Praat computer software package at 44100Hz. Each stimulus was read twice with the best example chosen by the second author[1]. Two normal children with no significant abnormalities in oral structures and functions, and who could speak clearly, underwent the same procedure to provide a control. The two speech therapists, who were responsible for 9 of the 10 CWAPD were interviewed about the language deficiency of each CWAPD. The language background of the participants is shown in Appendix B.

3.3. Measurement

The vowels were manually labelled according to acoustic and perceptual cues from F2 front to F1 terminal (Peterson et al., 1960). The duration was divided evenly into 10 sections, giving 11

[1] The second author chose the best example based on the judgment of the shortest acoustic distance between the performance of the CWAPD and the normal children in terms of F1 and F2.

data points. Following Cao (2007) on the duration ratio in the 1st and 2nd sonorous diphthongs and triphthongs, 6:4, 4:6, and 4:4:2 respectively, the method used for the monophthongs was applied to the diphthongs and triphthongs, getting each midpoint within a vowel. Several paired t-tests were run to compare the vowels articulated by each CWAPD and a normal child of the same gender.

4. Results

The results show that CWAPD have no difficulty pronouncing monophthongs as there were no significant differences between their vowels compared to those of normal children. However, the number of errors occurred with diphthongs and triphthongs. When a diphthong has [i] or [u] as its last segment, the least sonorant segment within the syllable, it is pronounced differently when compared to normal children. This means that the diphthongs /ia/, /ie/, /ua/, /uo/, and /ye/ are produced in a similar fashion to normal children. The diphthongs /iai/ and /iou/ did not induce any errors, meaning that /iau/, /uai/, and /uei/ are problematic. What remains of interest is that the relatively less sonorant segment was usually distorted whether there were two or three vowels.

In the following figures the blue lines represent the data from either Normal I or Normal II, depending on the gender of the CWAPD, while the red lines represent the children: A, B, C, D, E, F, G, H, I, and J. None of the participants showed any problems producing monophthongs and only the problematic diphthongs and triphthongs are reported.

4.1. Monophthongs

Children, even those with articulatory phonological disorders, acquire monophthongs earlier than diphthongs and triphthongs and we found no significant difference between the normal subjects and the CWAPD in their articulation of monophthongs. This agrees with what a number of scholars (Xu, 1987; Zhu et al., 2000a and 2000b) have observed in normal and abnormal children. As such, this study sheds light on the fact that CWAPD follow a similar vowel developmental process to that of normal children.

4.2. Diphthongs

In analyzing the diphthongs, the slope of each vowel's mid-point (F1 and F2) served as the dependent variable and was compared to either child I or II depending on the child's gender. In the production of [ai]: A had F1 of ($M=329.714$, $SD=119.568$), $t(6)=7.296$; B had F2 of ($M=1.295$, $SD=1362.686$), $t(6)=2.515$; C had F1 of ($M=235$, $SD=229.284$), $t(6)=2.712$, and with F2 of ($M=127$, $SD=134.284$), $t(6)=2.712$; H had F2 of ($M=854.429$, $SD=643.552$), $t(6)=3.513$; and I had F1 of ($M=-380.857$, $SD=305.233$), $t(6)=-3.301$, and F2 of ($M=736.714$, $SD=599.009$), $t(6)=3.254$, all $p<.05$. C and I showed difficulty moving upwards and forwards; B and H showed difficulty moving forwards; and only A showed difficulty moving upwards. For example, Figure 1 shows that C needs to move the tongue forward when producing the second segment, [i]. A

similar issue was found in the performances of B and H. I's [i] seemed to be deleted, while A's [i] was not produced as high as it should be. Overall, there was not a consistent pattern demonstrated by their production, but rather a distortion of the second vowel. To sum up, 5 out of the 10 children had problems producing [ai].

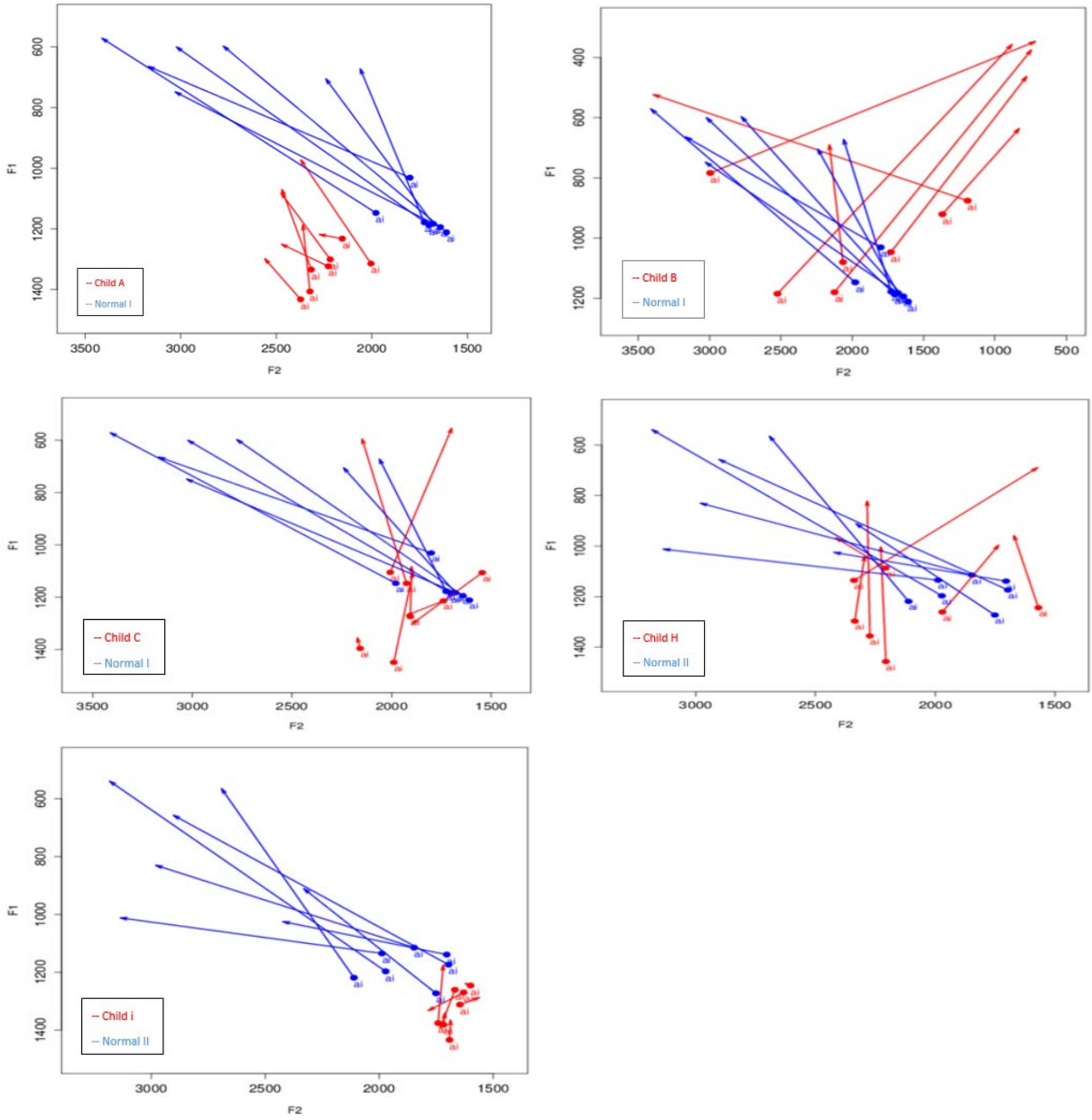


Figure 1: *[ai] produced by A, B, C, H, and I (left to right)*

In the production of [au]: B had F1 of (M=-326, SD=79.603); t(6)=-10.835, and D had F1 of (M=81.00, SD=52.898), t(3)=4.211, all $p<.05$. Both B and D had problems in moving upwards. However, Figure 2 indicates that D seemed to have problems with both [a] and [u], while B seems to move more upwards and backwards when producing [u].

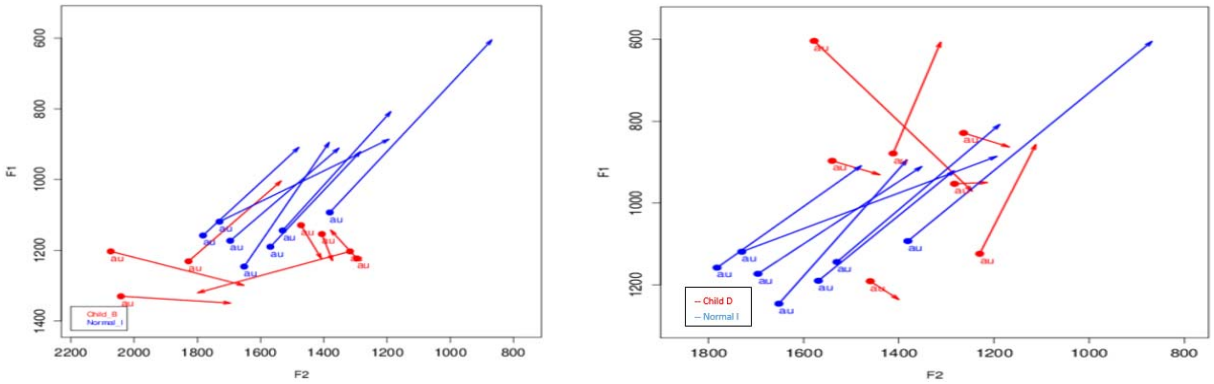


Figure 2: *[au] produced by B and D*

In the production of [ei]: F had F1 of (M=-220, SD=126.815), t(3)=-3.470, p<.05. This means F has a problem moving upward. However, Figure 3 shows that F seems to have problems with both [e] and [i].

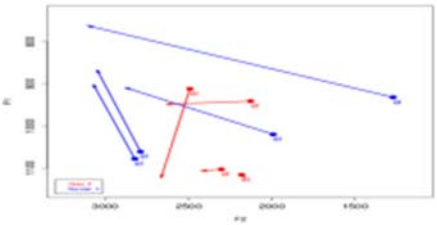
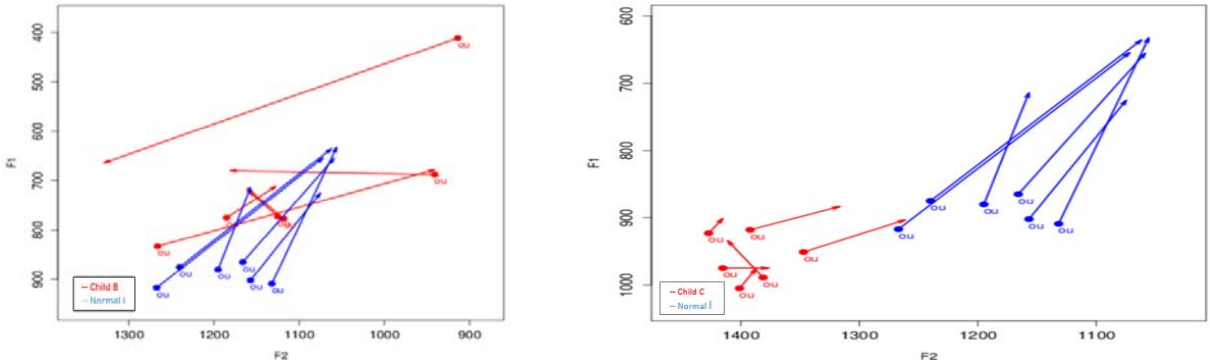


Figure 3: *[ei] produced by F*

In the production of [ou]: B had F1 of (M=-174.167, SD=92.244), t(5)=-4.625, and F2 of (M=-304.167, SD=283.03), t(5)=-2.632; C had F1 of (M=-124, SD=113.37); t(5)=-2.679, and F2 of (M=327, SD=104.384), t(6)=2.822; and child F had F1 of (M=-214.167, SD=123.543); t(5)=-4.246, and F2 of (M=-375.333, SD=244.979); t(5)=-3.753, all p<.05. This means that F had a problem in moving upwards and backwards, and B and C in moving upwards. However, C did not produce back vowels far enough back. Neither [o] nor [u] was produced as they should have been. Similar difficulties with [ou] were shown by B and F, as shown in Figure 4.



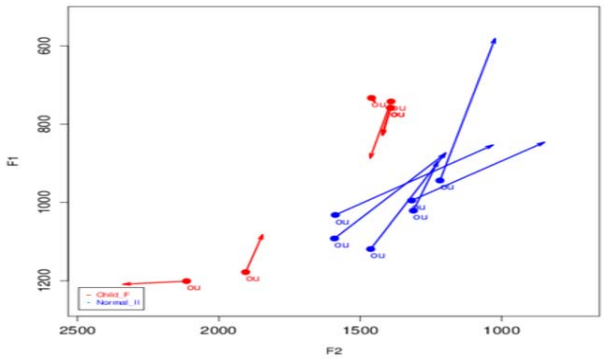
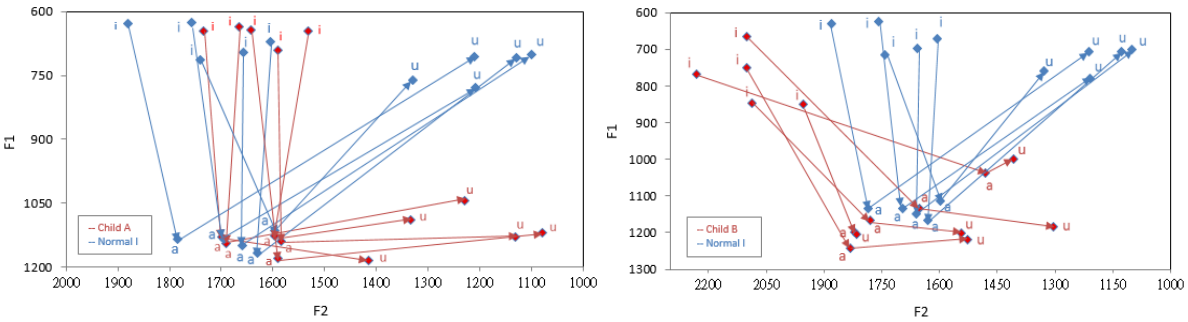


Figure 4: *[ou] produced by B, C, and F*

To sum up, the slope of the vowel only gives us information on dynamic movement, but a more detail description of the children's vowel performance is needed. It is hard to generalize the difficulty scale of diphthongs represented by our 10 CWAPD. It seems that [ai] is more problematic than other diphthongs—more than half the CWAPD made mistakes. However, we still find a pattern revealed by plotting F1/F2—some small changes need to be made to the articulation of the second vowel. Compared to other difficulties, the CWAPD seemed to be able to acquire [ai] earlier than other problematic diphthongs.

4.3. Triphthongs

In triphthongs, the slope of each vowel's mid-point (F1 and F2) served as the dependent variable and was compare to either I or II depending on the child's gender. This means that two separate paired t-tests covered the performance of one triphthong. Take [iau] for example: a paired t-test was run for [i] and [a] and another was run for [a] and [u]. In the production of [iau]: A in [au] had F1 of (M=-236.2, SD=151.828), t(2)=-5.340; B in [au] had F1 of (M=-750, SD=45.530), t(3)=3.967; and E in [au] had F1 of (M=-259.4, SD=172.657), t(5)=-3.359, all p<.05. All children had problems moving upwards in the last segment. Figure 5 demonstrates that A, B, and E need to practice more on the height of the last segment within the triphthong [iau].



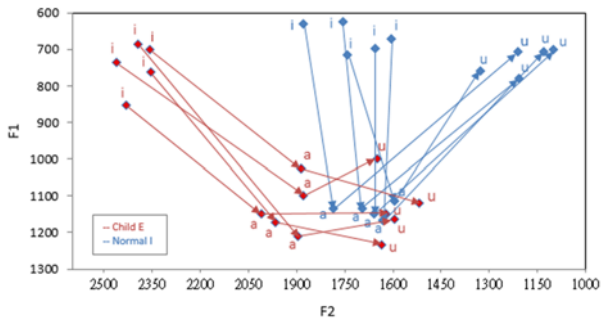


Figure 5: *[iau] produced by A, B, and E*

In the production of [uai]: A in [ai] had F1 of (M=-236.2, SD=166.62), t(4)=-3.479; B in [ai] had F1 of (M=227, SD=81.554), t(2)=4.821, and F2 of (M=-477, SD=92.065), t(2)=-8.974; C in [ai] had F1 of (M=-378, SD=43.715), t(2)=-14.977; and G in [ai] had F1 of (M=750, SD=45.530), t(3)=3.967, all p<.05. All children had problems moving upwards in the last segment, except B[2] who also had difficulty moving forwards. Figure 6 shows that A, B, C, and G need to practice more on the height of the last segment within the triphthong [uai].

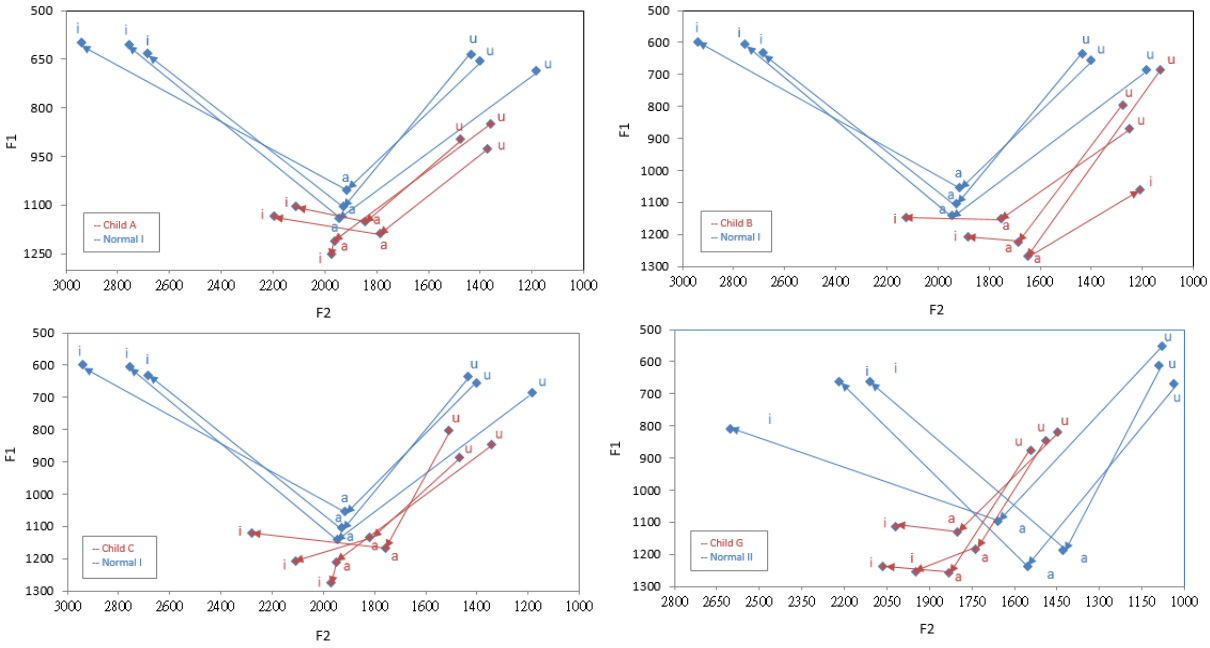


Figure 6: *[uai] produced by A, B, C, and G*

In the production of [uei]: C in [ei] had F1 of (M=-227.6, SD=155.523), t(4)=-3.272; F in [ei] had F1 of (M=750, SD=45.530), t(3)=3.967, and F2 of (M=1.318, SD=398.613), t(4)=7.392; G in [ei] had F2 of (M=1.165, SD=931.49), t(4)=2.796; H in [ei] had F1 of (M=-274.2, SD=200.344), t(4)=-3.060, and F2 of (M=1.447, SD=738.088), t(4)=4.384; I in [ei] had F2 of (M=610.6, SD=486.159), t(4)=2.808; and J in [ei] had F1 of (M=-274.2, SD=200.344), t(4)=-3.060, and F2 of (M=1.447, SD=330.083), t(4)=4.384, all p<.05. All children had problems in the last segment, in either horizontal or vertical dimensions. When producing the [i] of [uei], C needed to move his tongue upwards, and G and I needed to move their tongues forwards. F,

[2] Figure 6 shows that 1/3 of the repetition of B's [i] within [uai] was produced like [u]. It is hard to judge whether this is a mistake or not, but it caused a significant statistical difference in F2.

H, and J needed to move their tongues upwards and forwards. Figure 7 demonstrates that 6 out of the 10 children had difficulty producing [uai].

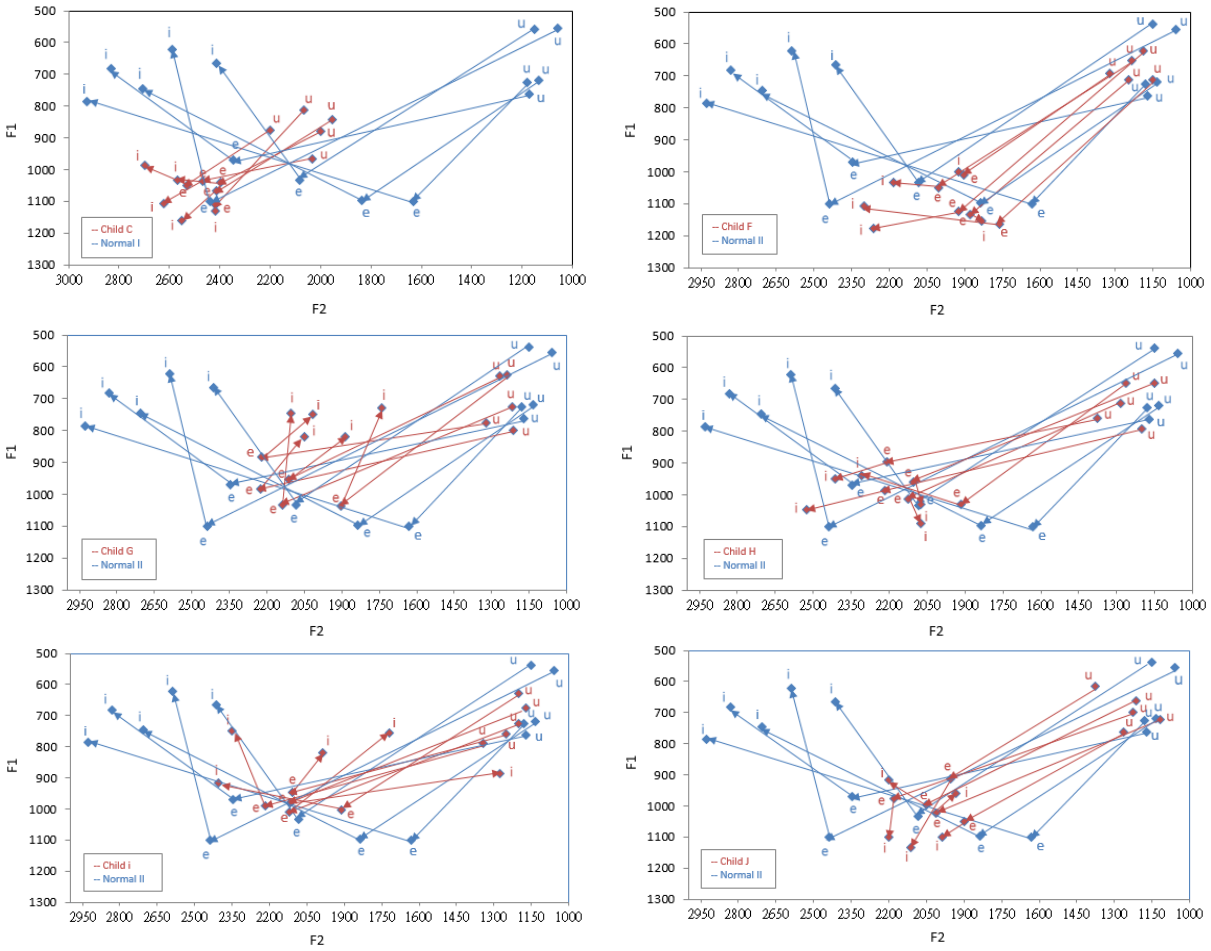


Figure 7: *[uei] produced by C, F, G, H, I, and J*

The slope of the vowel gives us some information, for example that the last segment is usually problematic. However, it is difficult to guide the participating CWAPD to move their tongues in the correct direction, either more forwards or backwards, solely based on the F1/F2 plotting for CWAPD generated in this study and we suggest that further examination is required. We could only identify the difficulty of a triphthong in its last segment, which is always the least sonorant segment within the syllable. In addition, it seems that [uei] was very problematic with more than half of the CWAPD making mistakes.

5. Discussion

Acoustic analysis of CWAPD from 3 to 5 years old has shown the order of acquisition of vowels from monophthongs to multiphthongs. This process parallels that of normal children—from a single vowel to several different ones within a syllable in Mandarin (Zhu et al., 2000a). In addition, the least salient sound within a syllable does not necessary cause errors, but rather when it is located in the last position, which is usually the least salient position. This means that the primary challenge that CWAPD face centers on the least salient components, both in terms

of the nature of the segment and its position. This study also observes that deviant pronunciations of CWAPD cannot easily be generalized, for example as to which dimension of the tongue needs more practice.

Distortion might adequately explain this phenomenon such as when a sound in a word is completely deleted, which is usually diagnosed as an omission by experienced therapists, for example [ue] for [uei] (喂). Distortion, however, where the sound is partially correctly produced, means some feature is distorted, such as [ɕan] for [san] (山). The difference between omission and distortion lies in the observation that the distortion patterns show clearly that the speakers have correctly acquired the metrical structure of diphthongs and triphthongs, i.e. they know that there are three vowels/moras involved. When the omission occurs, the metrical structure is not fully acquired, meaning the tone is likely not maintained. This study offers some acoustic data to show that CWAPD from 3 to 5 years old tend to distort rather than delete the least sonorous sounds in the last position, even though the therapists had not noted this. As such, emphasizing the distorted sound, i.e. encouraging CWAPD to perceive the difference between the correct sound and the distorted one, may be helpful. Once the diagnosis is correct, therapists can then use more methods to help CWAPD of 3 to 5 years old.

One may argue that omission or distortion of the sounds may provide an important distinction when correcting CWAPD. Our study has shown that different CWAPD may have different problems in producing the relevant sounds. Once one particular distorted sound has been detected, it is easier for therapists to emphasize this sound, for example, if a CWAPD distorts the sound [ai], as shown in Figure 1 with A, B, and I—A and B may distort [ai], while child I may delete [i] within [ai]. With child A, the tongue was not fronted enough when producing the [i] in [ai] and further practicing the movement from [a] to [i] is recommended. Making I perceive the [i] within [ai] may be the first priority. B, however, seems to confuse both [a] and [i] when producing the diphthong [ai], yielding a suspected sound of [au]. This suggests that practice combining both [a] and [i] is probably more helpful for him. As such, the contribution of this study is not only to highlight the difficulties shown by CWAPD from 3 to 5 years old, but also to provide acoustic evidence for theorists to help them choose the methods best suited to individual subjects.

Furthermore, this study aims to provide acoustic evidence to examine whether CWAPD reveal a pattern. Table 1 listed all the problematic diphthongs and triphthongs analyzed by our study. F, G, H, and I (all female) made fewer mistakes than A, B, C, D, and E (all male); the female CWAPD seemed to make mistakes in [uei], while only one male CWAPD did (C). Some mirror errors were found, e.g. A and B having problems with [iau] and [uai]. The relationship between the last two segments of a triphthong and a diphthong was also observed. For example, B could not produce [uai] and [iau] well, due to his difficulties articulating [ai] and [au]. A similar explanation may apply to A and C with [uai] and [ai], and F with [uei] and [ei]. One should cautiously note that this observation may not apply to all CWAPD from 3 to 5 years old: we found asymmetrical behaviors with H who could pronounce [uai] well, but not [ai].

The analysis of the therapists is shown with ours in Table 1. Our study indicates that: A's [iau], [uai], B's [ai], [ou], [uai]; C's [ou], [uai], [uei]; F's [ei]; G's [uai]; H's [ai]; and I's [uei] showed significant differences compared to those of normal children. Their therapists did not consider these sounds to be problematic. On the contrary, our study could not successfully

pinpoint C's [iau], G's [ai], and H's [ei], which were diagnosed as problematic sounds by their therapists. Generally speaking, our study indicates more problematic sounds than the therapists.

One might guess that the inconsistency lies in the fact that the production of CWAPD is not stable—when the CWAPD is under the supervision of their therapist, they may be able to pronounce the sound correctly. Once leaving the clinic, the CWAPD may forget how or not be required to pronounce the sound correctly, resulting in more errors in our study. However, this did not explain why some sounds diagnosed by therapists as problematic were not seen in our study.

Table 1: *I*[3]*. Comparison of error stimuli from our study and therapist diagnosis*

Child	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Our study	[ai], [iau], [uai]	[ai], [au], [ou], [iau], [uai]	[ai], [ou], [uai], [uei]	[au]	[iau]	[ei], [ou], [uei]	[uai], [uei]	[ai], [uei]	[ai], [uei]
Therapists' diagnosis	[ai]	[au], [iau]	[ai], [iau]	[au]	[iau]	[uei], [ou]	[ai], [uei]	[ei], [uei]	[ai]

6. Conclusion

To conclude, the least salient components in Mandarin for CWAPD aged 3 to 5 appear to be the greatest obstacle. This suggests monophthongs, which are more sonorant than multiphthongs, e.g. [a], are acquired first. Examining the syllable structure of Mandarin, only [i] and [u] occur in a triphthong's last position. As such, CWAPD may more easily improve their pronunciation by focusing on [i] or [u]. However, since more error types were found in diphthongs, and some CWAPD appeared to delete/distort the last segment, it seems that there is no consistent pattern regarding CWAPD performance. The individual distortion of sounds needs to be addressed one by one. It is possible that an individual CWAPD may have no problem producing [ai], but struggle with the [i] of [uai]. What concerns us is that last two segments [ai] and [uai] are identical. We might argue that this particular CWAPD could not articulate the [i] of [uai] because they were paying less attention to the final segment. The difference between the [i] of [ai] and [i] or [uai] lies in the fact that the latter structure is more complex than the former. Thus, it seems that our analysis provides a way to identify the location of errors, rather than offer an explanation.

On the other hand, regarding the difference between the researcher and therapist analysis, the therapist judged whether the sounds were correct in terms of the children's performance, while analysis with Praat dealt with acoustic signals, reflecting the trajectories of the tongue. One should note that in Peterson and Barney (1952) F1 values for vowels among 10 to 12-year-old English L1 children were as follows: [a] F1 values of 600–1200Hz; [i] and [u] F1 values of

[3] J's therapist did not take part in the interview. Since we needed to compare our data with the diagnosis data, we did not include her in this comparison.

300–500 Hz; and [e] F1 values of 650–750 Hz. The data produced by the CWAPD aged between 3 and 5 and with Taiwan Mandarin L1 in this study are: [a] F1 values of 1000-1400Hz; [i] F1 values of 400–1200 Hz; [u] F1 values of 600–1200 Hz; and [e] F1 values of 800-1150 Hz. Overall, the participating CWAPD from 3 to 5 years old had a larger F1 range in the vowels [i], [u], [o], and [a]—all F1 values were much higher than for normal children and 10 to 12-year-old English L1 children. It seems that children have difficulties in controlling the vocal organs to articulate all vowels at preschool age. One should also note that their F2 values ranges were similar to those of the normal children. This means that CWAPD from 3 to 5 years old may hyper-articulate the sound by over-rising or over-lowering the tongue, while controlling the tongue's extension seems not to be a problem. The results from analysis with Praat provides some details that describe the subtle differences made by CWAPD of 3 to 5 years old and this data can provide a reference to help therapists in their diagnoses.

References

American-Speech-Language Association (ASHA). 1993. *Definitions of Communication Disorders and Variations.* Avilable from www.asha.org/policy.

Bernthal, J., Bankson, N. W., & Flipsen, P., Jr. 2009. *Articulation and Phonological Disorders.* New York, NY: Pearson Higher Education.

Cao, Jian-Fen. 2007. ***現代語音研究與探索*** *[Modern phonetic research and exploration]*. Beijing: The Commercial Press.

Dodd, B., Leahy, J. and Hambly, G. 1989. Phonological disorders in children: underlying cognitive deficits. *British Journal of Developmental Psychology*, 7, 55-71.

Dodd, B. and So, L. 1994. The phonological abilities of hearing impaired Cantonese speaking children. *Journal of speech and Hearing Research, 37, 671-79.*

Fox, A. 1997. *Classification of speech disorders in German speaking children.* Unpublished MSc dissertation, University of Newcastle upon Tyne

Goldstein, B. 1996. Error groups in Spanish-speaking children with phonological disorders. In T. W. Powell (Ed.) *Pathologies of Speech and Language: Contributions of Clinical Phonetics and Linguistics* (New Orleans, LA: ICPLA), 171-177

Jeng, Jing-Yi. 2005.**不同言語速度、發語單位和發語位置對國語音段時長的影響** [The influence of speaking rate, utterance unit and position on segmental duration of Mandarin]. *Journal of National University of Tainan*, 39, 161-185.

Lin, Yen-hwei. 2007. *The Sounds of Chinese.* Cambridge University Press.

Lin, Yue-Xian. 2007. **學前輕度障礙兒童口語理解測驗之發展** [A developmental study on the auditory comprehension test for preschool children with mild impairment] Proceedings from *2007 National Science Council Sponsored Research in Early Childhood Education.* Taipei, Taiwan.

Lin, Bao-Gui, & Lin, Mei-Xiu. 1993. ***學前兒童語言障礙評量表之編訂及其相關因素研究*** *[*A study on the development of language abilities in preschool children in Taiwan*]* Department of Special Education National Taiwan Normal University.

Peterson, G. E., & Lehiste, L. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693-703.

Peterson, G. E., & Barney, H. L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.

Pickett, J. M. 1999. *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Allyn & Bacon.

So, L. and Dodd, B. 1994. Phonologically disordered Cantonese-speaking children. *Clinical Linguistics and Phonetics*, 8, 235-255.

Topbas, S. 1997. Phonological acquisition of Turkish children: implications for phonological disorders. *European Journal of Disorders of Communication*, 32, 377-396.

Van Riper, C. 1978. *Speech Correction: Principles and Methods* (6th ed.).Englewood Cliffs, N.J.: Pretice-Hall.

Wang, Nan-Mei, Fei, Pei-Ni, Huang, Xun, & Chen, Jing-Wen. 1984. 三歲至六歲學齡前兒童華語語音發展結構 [The Mandarin Chinese phonetic development structure for 3 to 6 year-old preschool children] *The Journal of Speech-Language-Hearing Association*, 2, 12-17.

Xu, Hong-Kun. 1987. **中國兒童學習國語及語法發展階段研究** [A study on the developmental stage of Chinese children's grammar learning in Taiwan] National Science Council. Taipei, Taiwan.

Zhang, Zheng-Fen, & Zhong, Yu-Mei. 1986. **學前兒童語言發展量表之修訂及其相關研究** [The Chinese version of preschool language scale and its related studies]. *Bulletin of Special Education*, 2. 37-52.

Zhu, Hua, and Dodd, Barbara. 2000a. The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(01), 3-42.

Zhu, Hua, and Dodd, Barbara. 2000b. Putonghua (modern standard Chinese)-speaking children with speech disorder. *Clinical Linguistics & Phonetics*, 14(03), 165-191.

Appendix A: 85 real Mandarin words

	i	**u**	**y**	**a**	**o**	**ɤ**	**ia**	**io**	**ie**	**ua**	**uo**	**ye**
zero initial	i	u	y	a	o	ɤ	ia	io	ie	ua	uo	ye
p	pi	pu	X	pa	po	X	X	X	pie	X	X	X
pʰ	pʰi	pʰu	X	pʰa	pʰo	X	X	X	pʰie	X	X	X
t	ti	tu	X	ta	X	tɤ	tia	X	tie	X	tuo	X
tʰ	tʰi	tʰu	X	tʰa	X	tʰɤ	X	X	tʰie	X	tʰuo	X
k	X	ku	X	ka	X	kɤ	X	X	X	kua	kuo	X
kʰ	X	kʰu	X	kʰa	X	kʰɤ	X	X	X	kʰua	kʰuo	X

	ai	**ei**	**au**	**ou**	**iai**	**iau**	**iou**	**uai**	**uei**
zero initial	ai	X	au	ou	iai	iau	iou	uai	uei
p	pai	pei	pau	X	X	piau	X	X	X
pʰ	pʰ ai	pʰ ei	pʰ au	pʰ ou	X	pʰ iau	X	X	X
t	tai	tei	tau	tou	X	tiau	tiou	X	tuei
tʰ	tʰai	X	tʰau	tʰou	X	tʰiau	X	X	tʰuei
k	kai	kei	kau	kou	X	X	X	kuai	kuei
kʰ	kʰai	X	kʰau	kʰou	X	X	X	kʰuai	kʰuei

Appendix B: Participants' background

Children Code	**gender**	**Birth**	**the date of collection**	**Language**	**Birthplace**
I	M	2011	27/07/2016	Taiwan Mandarin	Taipei
II	F	2011	23/12/2015	Taiwan Mandarin	Taipei
A	M	2010	30/10/2015	Taiwan Mandarin	Taipei
B	M	2010	04/01/2016	Taiwan Mandarin	Taipei
C	M	2010	08/12/2015	Taiwan Mandarin	Taipei
D	M	2010	27/10/2015	Taiwan Mandarin	Taipei
E	M	2010	13/11/2015	Taiwan Mandarin	Taipei
F	F	2011	18/12/2015	Taiwan Mandarin	Taipei
G	F	2011	02/12/2015	Taiwan Mandarin	Taipei
H	F	2012	01/12/2015	Taiwan Mandarin	Taipei
I	F	2012	19/10/2015	Taiwan Mandarin	Taipei
J	F	2011	27/10/2015	Taiwan Mandarin	Taipei

Therapist Code	**gender**	**the date of collection**	**occupation**	**Location**
X	M	01/06/2016	therapist	Taipei
Y	F	27/05/2016	therapist	Taipei

LingBaW *Linguistics Beyond And Within 4 (2018), 22-37*

Cognitive mechanisms and emergent grammatical features in Internet memes

Elke Diedrichsen
Institute of Technology, Blanchardstown, Dublin, Ireland

Abstract

Internet memes of the type composed of an image macro and text, have a strong form-meaning correlation that is shared among users of social media. Their frequency of usage and the immediacy of their broad reach around the world make them an interesting field of investigation for linguistic studies. I will argue in this article that Internet memes resemble linguistic signs. Users develop a literacy, i.e. a command of their usage through convention and shared usage history. Popular Internet memes can be found in a multiplicity of variations, where details of the shown picture are changed, while the general mood or topic of the meme, mostly expressed in the caption, remains broadly the same. This article will discuss cases where the development of meme variations works along the lines of known cognitive mechanisms like metaphor and metonymy, and their prerequisites, like abstraction. Some meme variations can be represented as grammaticalisation paths that lead to the emergence of grammatical features like morphemes.

Keywords: Internet memes, emergent grammar, grammaticalisation, metaphor, metonymy

1. Introduction: Memes are signs in human interaction

Richard Dawkins (1976) famously describes memes as the cultural equivalent of biological genes: A meme is a unit of human culture. It is comparable to a gene in that cultural units are passed on to other members of the culture. They may undergo variation, and there is competition among cultural units as some will be adopted and sustained within the culture, while others are not considered useful or interesting enough and will be forgotten by the next generation. Dawkin's examples of memes comprise pieces of music, linguistic theories, fashion trends, religious beliefs and ceremonies and the like. They also involve applied knowledge shared among humans, like knowledge about making pots, for example. There is a research area called 'memetics' that explores the insights to be gained from a comparison of cultural units and biological genes (see e.g. Blackmore 1999 and the contributions to the "Journal of Memetics" that appeared from 1999 through 2005 and is available online). Critiques of this approach maintain that cultural production and interaction cannot be easily broken down to

an exchange of "units" (Cannizzarro 2016, Shifman 2014). Producers of cultural content, which are human beings, should not be seen as mere 'carriers' of the content, but as wilful agents who select, alter and remove cultural material in their interactions and perceptions. This happens according to their free will, but also in relation to peer pressure, and requirements posed by the situational and general context.

I have, however, argued elsewhere (Diedrichsen 2013a, 2013b, forthcoming) that the term 'meme' is useful for the description of signs in human communication, as it opens up the explanation of the communicative potential of signs beyond the limiting boundary of the concept of a 'word', and in fact beyond the realm of mere written forms of linguistic expression. Any sign in human communication, of any shape, is a cultural unit. It emerges through usage (Wittgenstein 1960, Eco 1976, Feilke 1996, 1998, Everett 2012). The form and meaning of a sign is shaped and defined by usage conventions, which are acquired through interactions within a culture, and are therefore transient and subject to change. The same holds for structures and routines of linguistic expression. They are generated in interaction, as well. Before an individual interaction, structures and routinized norms may be known by interlocutors, but this is so because they have a "living historical continuity of interaction" (cf. Hopper 2011, Linell 1998: 59–60). So there are norms and structures in language, but they exist because interactants have come to know them through the interaction in a culture. Emergent grammar means that grammatical categories and structures do not freeze once they are established (Hopper 2011). They are, rather, being reinterpreted, modified and negotiated in every interaction. Note that this is the case for the shared knowledge base that underlies human communication as well: As Kecskes and Zhang (2009) point out, there is core and emergent common ground. While core common ground is a reliable source of mutually shared knowledge, some knowledge needs to be negotiated and confirmed in interaction. This is fleeting, unstable, situational knowledge that is not necessarily taken for granted by any of the interlocutors.

This paper will be concerned with Internet memes. Their appearance as a social media phenomenon, their conventionality and the regularities of their variation will be analysed from a linguistic point of view. The research questions underlying this analysis are the following: Can variations in Internet meme appearances be linked to cognitive mechanisms and to linguistic processes that lead to the emergence of new forms, like grammaticalisation? Can parts of memes consequently be described in terms of grammatical categories, like morphemes?

The paper will proceed as follows: Section 2 will define Internet memes and give a short characterization of their communicative potential. Section 3 will give an overview of the formal variation found in Internet memes and relate them to grammaticalisation processes.

2. Internet memes

Internet memes are signs that follow conventions. They are used for communication. It has been argued that the interaction with Internet memes should be described as *phatic*: In phatic communication, it is the participation that counts, rather than the delivery of a message (Milner 2012).

Internet memes are contentful objects of diverse shapes that are spread across Social Media in order to attract viewers and invite them to produce related content, mostly with the intention of being funny or making a comment about an aspect of life in general or popular culture. Meme usage has aspects of communicative interaction, because memes have established conventions of usage. They are produced for an audience in order to spark reactions, and they are viewed against the background of a culture of meme usage.

Each meme has its own topic, refers to a certain circumstance, mood or attitude that other users can relate to. Within this broad topic area, users are free to create variations and combinations of memes. Digital communities have "sets of unwritten rules for proper meme-related conduct" (Katz and Shifman 2017: 828). The site knowyourmeme.com serves as a reference that formulates current proper usage of memes. "Becoming literate in these groups is a phatic process, since levels of meme literacy serve as cultural capital, differentiating members from non-members." (Katz and Shifman 2017: 828, see also Nissenbaum and Shifman 2015, Miltner 2014).

The aspect of global phatic communication via the Internet, the experience of making communicative interaction possible in real time with multiple participants around the world, and interactively inventing signs, rules and conventions, has led me (Diedrichsen forthcoming) to the following definition of memes, which is a slight variation from the definition offered by Shifman (2014: 41):

One meme is:

a) A group of digital items which share common characteristics of content, form, and/or stance
b) Users create these items with the intention to make them interact with and react to each other
c) The items are circulated, imitated, and/or transformed via the Internet by many users.

The semantics of memes is mainly informed by participating users' relevant cultural knowledge (interests relating to age group, pop culture, country and culture of origin, familiarity with technology and internet, relation to sentiments shared by users, like the love of cats or the appreciation of certain films and their characters). The form of memes is mainly shaped by the technically feasible, with respect to the publishing source and the capability and motivation of users to generate content, but also by trends and existing forms of memes. An important aspect in the creation of memes is also the desire to attract attention, which is achieved by the choice of topic, use of colours, recognisable image structures, bold letters, and the size of the image in this digital linguistic landscape (Diedrichsen forthcoming).

3. A arammar of Internet memes?

Internet memes are a very popular and observable example of how signs and sign variations emerge through usage. They are created in real time by Internet users, and any new creation is attested on the Internet. The Internet is a medium where the creation of content leads to immediate mass distribution, and the extent to which it is received is unpredictable. Also, any reaction to the content can be immediate. There is a steady stream of new picture- and text combinations that may become memes if they catch on with users and attract many variations. The form of the meme is varied and generally unpredictable. A variety of meme genres have

been listed and explained by Shifman (2014: 108-115). Any list of memes will be preliminary, however, because memes keep emerging, and taking on new shapes, such that there cannot be any definite description of the form a meme can take. I will limit the discussion in this paper to one very popular meme genre, that is called *stock character macros* or Advice Animals. These are images that show a character or animal on a coloured background with a caption in white bold text. The descriptions of these memes are available via a database search for the meme name in knowyourmeme.com.

I will argue in this paper that the variations undergone by memes show features that are reminiscent of cognitive mechanisms related to grammaticalisation processes and emergent grammatical categories in language.

3.1. Grammaticalisation: The emergence of grammatical categories

Grammaticalisation is a process by which linguistic elements with no or little grammatical function obtain more grammatical function or the grammatical function gets established in the first place. The process of the establishment of a grammatical function frequently happens through reanalysis, which is a metonymical process. A given structure obtains a new interpretation in terms of the functions of its elements. This happens in circumstances where the interpretation of the structure is ambiguous. The result of a reanalysis is a new interpretation of the given structure. This new interpretation is not obvious on the structure, unless it is applied to variations of the structure. The re-application of the new interpretation is called analogy, and analogy is guided by metaphor. Analogy makes the changes that have taken place in reanalysis observable. Reanalysis and analogy are the major mechanisms in language change, and grammaticalisation does not occur without them (Hopper and Traugott 1993: 61–62). Metaphor and metonymy are mechanisms used by human cognition for problem solving.

> 'The main direction of both types of problem solving is toward informativeness, but the two types correlate with shifts along different axes. Metaphorical change involves specifying one, usually more complex, thing in terms of another not present in the context. Metonymic change, on the other hand, involves specifying one meaning in terms of another that is present, even if only covertly, in the context. It is largely correlated with shifts to meanings situated in the subjective belief state or attitude toward the situation, including the linguistic one. While metaphor is correlated primarily with solving the problem of representation, metonymy and lexicalizing of conversational meanings are correlated with solving the problem of expressing speaker attitudes.'
>
> (Hopper and Traugott 1993: 87).

One famous example of the effect that can be brought about by reanalysis and analogy is the emergence of the English *going to* future. It will be explained here shortly, and after that I will proceed to discuss the application of the terms reanalysis and analogy to processes going on with memes. The discussion will lead me to the conclusion that the cognitive mechanisms metaphor and metonymy apply to the description of meme variation, and that in some cases one can identify grammaticalisation processes that result in the emergence of morphological and syntactic categories.

3.2. The grammaticalisation process of the English going to *future*

The grammaticalisation of the English *going to* future comes about because there is a future component existent in the interpretation of *going to* even in its literate, full verb reading: A statement about going somewhere involves that the arrival at that location will be in the future.

(1) a. *I am going to London.*
b. *I am going to visit Bill.*
c. I am going to like Bill.
d. I am gonna like Bill.

If I say the sentence in (1a), I will arrive in London in the future. Here, the full verb use of *going to* is obvious because there is no other verb in the sentence. The full verb use is less obvious if the specification of the destination involves another verb, as in (1b). Here, it is possible to understand the sentence in a similar sense as (1a), which is 'movement to a place where Bill lives', which implies that the speaker will see Bill in the future. The fact that there are two verbs in the sentence, however, might lead the interpreter to a reading where *going to* is a future auxiliary and *visit Bill* is the intention projected into the future. Auxiliaries exist in English, therefore this assumption would not require the invention of a whole new grammatical category. The ambiguity of the verb function is the first stage in the grammaticalisation process. In the future reading, a reanalysis of the structure has taken place, where *going to* is not a motion verb and *visit Bill* is not a place, which would be the interpretation to be gained from a full verb reading. The fact that the reanalysis has taken place is, however, not visible in the structure. It becomes obvious when the new structural interpretation is applied to new uses of *be going to*, where the interpretation of 'motion to a place' is not possible, as in (1c). In a further step, the complex auxiliary *be going to* becomes a single morpheme, *gonna*, by phonetic reduction. According to Hopper and Traugott (1993:61), the process behind this is again reanalysis.

Table 1: *The grammaticalisation path of the going to future (after Hopper and Traugott 1993: 61)*

Grammaticalisation stage	Perceived Structure	Mechanism for interpretation
Stage 0	I am going [to London]	Full verb reading of going: motion verb To London: Specification of destination
Stage 1	I am going [to visit Bill]	Ambiguous structure with two verbs. Full verb reading of going: motion verb In analogy with stage 0, the purposive clause to visit Bill is interpreted as a specification of destination.
Stage 2	I am [going to] visit Bill	Reanalysis of the ambiguous structure: Going to: Future auxiliary Visit: Activity verb,
Stage 3	I am [going to] like Bill	Analogy: going to as a future auxiliary is used with a stative verb, which is possible, as the auxiliary reading of going to does not require any implication of change of place
Stage 4	I am [gonna] like/visit Bill	Reanalysis: The complex auxiliary is reduced to a single morpheme; any ambiguity is removed.

4. Emergent grammar and grammaticalisation processes in Internet memes

This section will introduce a few cases of grammaticalisation in terms of emergent grammatical categories in memes. It can be shown that the frequency of usage and the multiplicity of variations that the memes are subjected to bring about shifts in the functionality of pictorial elements that can be described along the lines of grammaticalisation processes as outlined in section 3.

The viral nature of the memes involves their recognisability, but some variation as well, as they are applied to new stories within the same broad topic area. In the creation of new meme variations, it happens that some formal elements are retained and others are altered in a generally straightforward fashion. Here, I will mainly talk about changes that resemble the emergence of new morphemes, but it will be argued at the end of the paper that instances of syntactic organisation can be found in meme usage as well.

A reviewer remarks that the formal variation in memes should not be described as grammaticalisation processes involving, for example, morphemes, but rather as developments that resemble grammaticalisation processes with features that resemble morphemes. I have made the point here that Internet memes can be treated analogous to linguistic signs because of their communicative nature and their conventionality. I am arguing that like linguistic signs, Internet memes are cultural units that evolve and spread among a community of users, and that their usage and development are subject to communicative demands. With that analogy in mind, I am proceeding to regard formal developments and shifts in interpretation along the lines of grammaticalisation processes. This involves the assumption that in some cases the emergence of units carrying meaning and forming new complexes can be described as grammaticalisation processes leading to the formation of morphemes. Note that morphology and the description of morphemes is not restricted to spoken and written words in linguistics. Sign languages that are visual-gestural and have no spoken or written forms do have rich inflectional morphology in terms of hand shapes, hand movement, facial expressions and other visible articulators (Leeson and Saeed 2012). It is with regard to this broader view of communicative means that I describe the features and processes analysed here, using linguistic terminology for a phenomenon that I consider to be a linguistic one, but in a broader sense that extends beyond the realm of words and sentences.

4.1. Meme: Roll Safe

This meme was very popular around the beginning of the year 2017. It involves a picture of a character called Roll Safe from the web series *Hood Documentary,* smirking and pointing his finger at his temple. The text jokingly points out a kind of reasoning where a circumstance is avoided by removing its necessary and sufficient condition (NSC). This meme has undergone some variation, but certain elements of the meme remained unaltered. Viewing memes and their instantiations as complex *signs,* as pointed out above, I give an overview of their features in terms of the main areas of linguistic description, morphology, semantics, and pragmatics. The semantics part describes the meaning and propositional content of the meme, as it has been conventionalised over manifold instances of usage, and as it is described in

www.knowyourmeme.com. The pragmatics part involves the communicative function that can be associated with the meme. The morphology part works on the claim made here that morphemes emerge and can be isolated in the analysis of memes. Note that the discovery of morphemes involves an analytical observation of a variety of instances of a meme, such that it becomes obvious which is the part of the meme that gets transferred to other variations of the meme, and carrying semantics in the course of this transferral. I will provide a constructional schema for every meme discussed here, that informs about these features. Note that an important area of linguistic investigation, syntax, is left out here, because we discuss memes as singular items. The possibility of analysing combinations of meme contents will be left for future work.

Table 2: *Constructional Schema for Roll Safe meme*

Meme: Roll Safe	Morphology	Morpheme: 'Finger-to-temple' gesture The same gesture is used in different variations of the meme.
	Semantics	Joke relating to a kind of reasoning where a circumstance is avoided by removing its condition. The gesture signifies that reasoning is involved.
	Pragmatics	Irony, mockery: The reasoning portrayed in the meme is stupid and nonsensical, applies to 'losers'

The images in Figures 1–5 show instances of the Roll Safe meme. We need to look at a few instances in order to see what variations occur. Comparing the images, it becomes clear that the caption expresses variations on an ironic reference to a kind of reasoning where a circumstance is avoided by removing its NSC. The characters used in the meme change, but what remains constant is the finger-to-temple gesture. Compare the following images:

Figure 1: *Roll Safe Job (http://thinkingmeme.com/top-24-roll-safe-meme/ Last accessed 17 January 2018)*

Figure 2: *Roll Safe Walter White (http://thinkingmeme.com/top-24-roll-safe-meme/ Last accessed 17 January 2018)*

Figure 3: *Roll Safe Batman (http://thinkingmeme.com/top-24-roll-safe-meme/ Last accessed 17 January 2018)*

Figure 4: *Roll Safe Trump (http://thinkingmeme.com/top-24-roll-safe-meme/ Last accessed 17 January 2018)*

Figure 5: *Roll Safe balance (http://www.ebaumsworld.com/pictures/14-of-the-best-roll-safe-memes/85257795/ Last accessed 17 January 2018)*

From the observation that the pointing gesture reappears in variations of the meme, carrying the meme semantics, we conclude that this gesture is acting like a morpheme. Note that the usage of this gesture as main contributor to the expressive power of the meme was predictable, as there is a convention for the same interpretation of this pointing gesture available in our culture. The gesture is a known and well used metaphor for a certain mindset that involves reasoning, and it can be used either seriously or ironically. From the original image featuring Roll Safe and its gesture, the grammaticalisation path towards the morphological function of the gesture can be outlined as in Table 3.

Table 3: *Grammaticalisation path for Roll Safe meme*

Grammaticalisation stage	**Picture**	**Perceived Structure**	**Mechanism for interpretation**
Stage 1	Figure 1	[Character called Roll Safe from web series *Hood Documentary* smirking and pointing his finger at his temple]	Full image carries meme semantics
Stage 2	Figure 1	Character called Roll Safe from web series *Hood Documentary* smirking and [pointing his finger at his temple]	Reanalysis [metonymical]: The meme semantics lies in the pointing gesture, the character could be replaced. This interpretation is not random, as there is an existing meaning convention for the pointing gesture
Stage 3	Figure 2–5	Any character pointing their finger at the temple	Analogy [metaphorical]: The pointing gesture can be used on any character, and carries the meme semantics by analogy

4.2. Meme: Scumbag Steve

Scumbag Steve is a meme that is very well known and can almost be considered a classic within meme/popular culture. The meme started spreading and peaked in the year 2011, according to Google Trends and knowyourmeme.com, but it is still widely known and appears in social media from time to time. The photo displayed in the meme shows a young man with an iconic brown cap worn sideways, a golden chain around his neck and an open brown coat with a furry hood, who stands in a doorframe. The text shown in the meme characterises this person as someone who shows mean, antisocial behaviour. In the course of the usage of this meme, variations occur that explain different situations where someone has shown antisocial

behaviour. Interestingly, some variations also alter the composition of the picture in that one element is singled out to carry the 'scumbag' semantics to alternative situations. It turns out that the element that is interpreted as the carrier of the 'scumbag' semantics is the hat, and it may be placed elsewhere, for example on objects, in order to express that these are malfunctioning. The process of reinterpreting the whole meme in such a way that the hat becomes the main carrier of the 'scumbag' semantics is a double metonymic expansion as described by Ruiz de Mendoza and Galera (2014: 119–123): The hat stands for the wearer, and, in terms of the AGENT FOR ACTION metonymy, for the behaviour of the wearer. As there is no existing convention for a brown cap to signify mean behaviour, this grammatical feature can be said to have emerged through usage.

Table 4 shows the Constructional Schema of the meme, that explains its semantic, morphological and pragmatic features, based on observations made in a picture selection as it is exemplified in Figures 6–11. There are variations of Scumbag Steve featuring an empty doorway, that imply that Scumbag Steve is gone when you need him (cf. Figure 12). I have not found any instance of a 'scumbag' meme featuring one of the other details seen in the Scumbag Steve meme, like a furry coat, or a golden necklace.

Figure 6: *Scumbag Steve iPod stolen (http://www.lolriot.com/scumbag-steve-compilation-20-images/scumbag-steve-invite-steve-to-party-ipod-stolen/ Last accessed 23 January 2018)*

Figure 7: *Scumbag Steve beer (https://goo.gl/images/tK5ebR Last accessed 23 January 2018)*

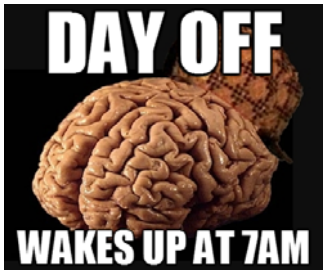


Figure 8: *Scumbag Brain (https://goo.gl/images/XF88jE Last accessed 23 January 2018)*

Figure 9: *Scumbag Printer (https://goo.gl/images/eB683t Last accessed 23 January 2018)*

Figure 10: *Scumbag Hot Pocket (https://goo.gl/images/ei1REN Last accessed 23 January 2018)*

Figure 11: *Scumbag ceiling fan (https://goo.gl/images/bPjaTu Last accessed 23 January 2018)*

Figure 12: *Scumbag Steve Doorway (http://i0.kym-cdn.com/photos/images/original/000/545/804/775.jpg Last accessed 23 January 2018)*

Table 4: *Constructional schema for Scumbag Steve meme*

Meme: Scumbag Steve / Scumbag Hat	Morphology	Morpheme: brown cap The cap is used in different variations of the meme, and it is used independently in a meme of its own, to signify mean behaviour of objects
	Semantics	Person behaving mean and egoistic, antisocial, also: malfunctioning objects
	Pragmatics	Complaint

From the original image featuring the young man called Scumbag Steve, the grammaticalisation path towards the morphological function of the hat can be outlined as shown in Table 5. Figure 12 does not appear on this grammaticalisation path, as it is a deviation from the general tendency to let the hat appear in images to carry the Scumbag Steve semantics. It demonstrates, however, that the choice of the hat is indeed random, and the detail singled out could have been anything in the original picture, like the doorway, for example. Because of the emergent nature of meme creation, this may certainly still happen. For the majority of variations of the Scumbag Steve meme, it is currently the hat that ascribes 'scumbag' behaviour to everyone and everything one wants, or rather, everything that can be pictured. This fact has been noticed by users and made into a meme (Figure 13).

Figure 13: *Scumbag Meta comment (https://goo.gl/images/LfPjEf, last accessed 23 January 2018)*

Table 5: *Scumbag Steve/Scumbag Hat: Grammaticalisation path*

Grammaticalisation stage	Picture	Perceived Structure	Mechanism for interpretation
Stage 1	Figure 6–7	[Young man with a furry coat, a golden necklace and a brown cap standing in a doorway]	Full image carries meme semantics
Stage 2	Figure 6–7	Young man with a furry coat, a golden necklace and [a brown cap] standing in a doorway	Reanalysis [metonymical]: The meme semantics lies in the brown cap, the character and other details could be replaced/discarded. This interpretation is random: There is no existing meaning convention for the brown cap
Stage 3	Figure 8–11	Any character or object with the brown cap	Analogy [metaphorical]: The brown cap can be used on any character or object, and carries the meme semantics
Stage 4	Figure 8–11	The brown cap as a meme of its own (cf. knowyourmeme.com)	Reanalysis: The brown cap on its own is "Scumbag Hat", the complex 'Scumbag Steve' meme is reduced to a simple item that acts as a morpheme (cf. Hopper and Traugott 1993: 61, where the reduction from *going to* to *gonna* is described as reanalysis as well)

4.3. Meme: Arthur's Fist

We have noted in the discussion of the Scumbag Steve meme that the abundance of variations on memes may involve creations that make a meta-comment about the meme itself, in terms of ridiculing its content or the way it is displayed. One meme whose instantiations show such meta comments quite frequently is the Arthur's Fist meme. Arthur's Fist shows a screenshot from the children's television series called "Arthur", in which the protagonist animated character clenches his right fist, while wearing a yellow sweater and blue trousers.

In the meme, this image appears with captions that describe situations in which the user suppresses disappointment or other negative emotions. The image was first used as a meme in late summer 2016, and its search interest peaked between July and October 2016 (knowyourmeme.com, Google Trends). Arthur's Fist is not easily described in terms of emergent morphology. It turns out that the many meta variations that occur with this meme do not single out one element only. The obvious candidate for a morpheme carrying the semantics of the whole meme would be the fist itself, as it is the most prominent element in the picture. Also, a clenched fist is a known indicator of suppressed emotions. It is an indexical sign. In the meme, it has a metonymical relationship towards the kind of emotion that it is supposed to express. Metonymies play an important role in structuring emotion categories (Ungerer and Schmid 1996: 128). The attempt at keeping a negative emotion under control is part of the emotion scenarios provided by Lakoff 1987: 377ff. and Kövecses 1990, cf. also Ungerer and Schmid 1996: 141), and its visible physical symptom is the clenched fist.

In the Arthur's Fist meme, the relationship between signifier and signified does not emerge through convention, as it is the case with Scumbag Hat, for example. Rather, users know from experience that a clenched fist can be part of a frustrating situation, and they know from media experience that the display of a clenched fist can signify suppressed frustration by metonymy, like a red, steaming head, shown in a comic, for example, can signify anger.

In variations of the meme, multiple elements of the image undergo variations. The meme appears with a clenched left fist, with fists belonging to other characters known from television series or movies, like Spongebob or Hulk, the fist may have a dark skin colour, or it may not be seen because the arms of the sweater are too long (cf. knowyourmeme.com and Google images for samples of the meme). The caption in these images generally expresses frustration, which bears a relation to the way the image is altered. There are also instances of the Arthur's Fist meme where the modality of the image representation changes, as shown in the examples below. From my observation, an unusually high number of variations of the Arthur's Fist meme acts as a meta-comment, in that the caption makes a comment that is reflected in the picture. It seems that with this meme, phatic communication in the sense of displaying insider knowledge, playful variation of and reaction to a picture, is realised to a very strong degree. Many instances of the meme (cf. Figures 16–19) are self-referential, in that their main function is a playful variation of the picture components, rather than a comment about a state of affairs in the world. Many memes do not 'make sense' in terms of referring to the outside world, but rather represent jokes in and by themselves (see Katz and Shifman 2017 for a discussion and examples).

The constructional schema in Table 6 shows the formal, semantic and pragmatic properties of the meme Arthur's Fist, based on observations made in a picture selection as it is exemplified in Figures 14–19.

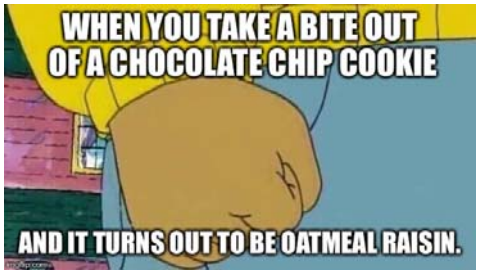


Figure 14: *Arthur's Fist Oatmeal Raisin (https://imgflip.com/i/1l0bde Last accessed 20 December 2017)*

Figure 15: *Arthur's Fist charger (https://goo.gl/images/jBujU9 Last accessed 28 January 2018)*

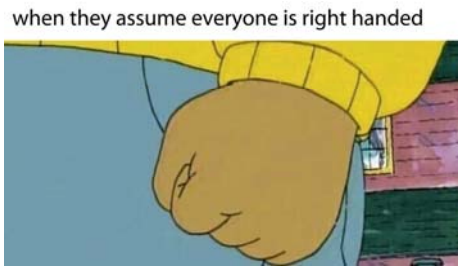


Figure 16: *Arthur's Fist left handed (https://goo.gl/images/t5q8bT Last accessed 28 January 2018)*

Figure 17: *Arthur's Fist sweater (https://goo.gl/images/vVRx1N Last accessed 28 January 2018)*

Figure 18: *Arthur's Fist lego (https://goo.gl/images/LMhjDo Last accessed 28 January 2018)*

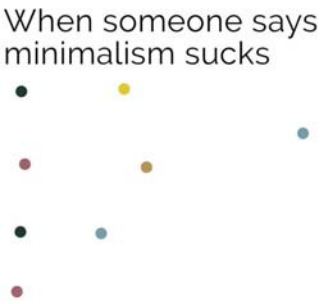


Figure 19: *Arthur's Fist minimalism (https://goo.gl/images/c3SgMT Last accessed 28 January 2018)*

Table 6: *Constructional schema for Arthur's Fist meme*

Meme: Arthur's Fist	Morphology	Clenched fist in yellow sweater Colour combination of yellow, blue and brown-ish (skin-colour); green and pink in the background. These components reappear across replications of the meme
	Semantics	Something happens that speaker does not like, and their helpless response is a clench of the fist.
	Pragmatics	Complaint, expression of helplessness, often: meta comment, play with components and modality of the picture, phatic communication

In the variations on Arthur's Fist meme, it is not easy to identify a direction of development. The instances of the meme show variations not only on disappointing situations in the world that one might suffer, but also on the topic 'fist' itself, and the mode of image representation found in the meme. In the variation schema presented in Table 7, that is not quite a grammaticalisation path, three lines of development are distinguished: The first one describes the statement the meme makes about disappointments in the world, the second one sees the 'fist' as a general topic, detached from the 'Arthur' context, that is subject to playful comments that involve possible circumstances for fists, like handedness, sweaters' arm length, or, not pictured here, the colour of the individual's skin, or the recognition that the fist belongs to another character like Hulk or Spongebob. One can argue here that a topic-comment structure is established with the fist as the recurrent topic and the changing colour, shape or other circumstance around it as the comment, the new information. In grammatical terms, this would resemble a subject-predicate structure, as the predicate ascribes a feature to the subject.

The third line of development goes towards abstraction, one of the formal cognitive operations described by Ruiz de Mendoza and Galera (2014: 86–90). Both metaphor and metonymy are based on abstraction. A formal cognitive operation is a mental mechanism by which language users can access, select, abstract, substitute, and integrate conceptual structure for the production and interpretation of linguistic (or other, my addition) communicative content (Ruiz de Mendoza and Galera 2014: 85). With the Arthur's Fist meme, the abstraction involves an iconic, i.e. metaphorical mapping of the image to different modes of representation, including, for example the layout of the scene in Lego bricks or the deconstruction of the image into the minimalist colourful dots.

Table 7: *Three lines of development for Arthur's Fist meme*

Grammaticalisation stage	**Picture**	**Perceived Structure**	**Mechanism for interpretation**
Stage 1	Figures 14–15	Drawn image of clenched fist in yellow sweater, blue trousers and pink/brown background are visible	Known Metonymy: Clenched fist as symptom of an attempt to control a negative emotion
Stage 2	Figures 16–17	Variation on representation of known fist image: The fist is isolated and introduced as topic	Focus on fist as a topic. Related matters like handedness, sweaters' arm length etc. are offered as comments. Topic-Focus arrangement, resembles syntactic subject-predicate structure
Stage 3	Figures 18–19	Variation on representation of known fist image: Changes in the modality of picture representation	Iconic mapping of the image to different modes of representation [Metaphor]

5. Summary and conclusion

In this paper, the popular communication form of Internet memes has been analysed in terms of their resemblance with linguistic units. The considerations started out from the observation confirmed in the literature that Internet memes are not just visual jokes dotted around the net, but that they have communicative functions, they have sign value, and they follow conventions. For each meme, a general topic area is identifiable that the picture/text combination plays with and builds up on. Through frequent usage and free creative variation, meaning elements remain relatively stable. The formal variation is generally unlimited, but often, one detail is kept throughout all variations, in order to make the meme recognisable. It has been shown that these consistent formal details behave like grammatical features, i.e. morphemes, as they carry semantics known from the full meme complex. Because the variation of memes is an undirected interactive process, the development of the grammatical features is not generally predictable. The paper discusses three popular Internet memes: Roll Safe, Scumbag Steve and Arthur's Fist. The Roll Safe meme includes a pointing gesture that can be shown to develop into a free morpheme, as it combines with new characters for variations of the meme. This was predictable as there is already a meaning convention for this gesture. In variations of the Scumbag Steve meme, the brown cap worn by the character develops into a free morpheme that combines with new characters and objects for variations of the meme. This was not predictable, as there is no

existing convention for a brown cap to signify mean behaviour. Both the cap as a morpheme and the behaviour attributed to it by metonymy emerge from the interaction with this meme alone. With the meme Arthur's Fist, the line of development is different from the first two memes discussed. The fist is a prominent component in the picture, and it is conceivable that it could develop into a morpheme-like item, but this is not what happens in the majority of variations of the meme. Rather, there are two lines of development that both involve meta-commentary on the meme: The first one introduces the fist as a topic, and gives a variety of comments on it, involving handedness, skin colour, the visibility of the fist and the like. I have argued that these instances where the fist is singled out from the correlation with the meme and taken as a topic on its own bring about topic-comment structures, which are the functional equivalent of subject-predicate structures. As for the other line of development, there are many instances that show abstractions of the image, its components and its colour scheme, by mapping it to a different mode of representation, i.e. a layout in Lego bricks or a minimalist dotted picture with the recognizable colours of the original image. For all memes discussed here, I have argued that the variations shown can be described in terms of cognitive mechanisms like metaphor and metonymy, which are operations that have been associated with grammatical changes in linguistic theory. For the first two memes, the changes can be described in terms of grammaticalisation paths, where reanalysis of a known structure leads to the development of a new morpheme. For Arthur's Fist, metonymy and metaphor can be identified as mechanisms for interpretation, but the variation in the meme does not lead to an identifiable grammaticalisation path. Given that the structures that emerge in communicative interaction are never predictable, this last example demonstrates the freedom that generally exists for interactive patterns to emerge. The objective of the approach presented here is to support a discussion that extends the description of emerging grammatical features to communicative means beyond language

References

Blackmore, S. (1999). *The Meme Machine.* Oxford: Oxford University Press.

Cannizzarro, S. (2016). Internet Memes as Internet Signs: A Semiotic View of Digital Culture. *Sign Systems Studies* 44(4), 562–586.

Dawkins, R. (1976). *The Selfish Gene.* Oxford: Oxford University Press.

Diedrichsen, E. (2013a). Constructions as Memes – Interactional Function as Cultural Convention Beyond the Words. In Liedtke, F. & Schulze, C. (Eds.): *Beyond Words* (pp. 283–305). Berlin: De Gruyter.

Diedrichsen, E. (2013b). From Idioms to Sentence Structures and Beyond: The Theoretical Scope of the Concept 'Construction'. In Nolan, B. & Diedrichsen, E. (Eds.): *Linking Constructions into Functional Linguistics – The Role of Constructions in Grammars* (pp. 295–330). Amsterdam: John Benjamins.

Diedrichsen, E. (forthcoming). On the Interaction of Core and Emergent Common Ground in Internet Memes. To appear in *Internet Pragmatics,* special issue on the Pragmatics of Internet Memes.

Eco, U. (1976). *A theory of semiotics.* Bloomington, London: Indiana University Press.

Everett, D. (2012). *Language, the cultural tool.* London: Profile Books.

Feilke, H. (1996). *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik,* Frankfurt am Main: Suhrkamp.

Feilke, H. (1998). Idiomatische Prägung. In Barz, I. & Öhlschläger, G. (Eds.): *Zwischen Grammatik und Lexikon* (pp. 69-80), Tübingen: Max Niemeyer.

Hopper, P. J. (2011). Emergent Grammar and Temporality in Interactional Linguistics. In Auer, P. & Pfänder, S. (Eds.): *Constructions: Emerging and Emergent* (pp. 22–44) Berlin/New York: De Gruyter.

Hopper, P. J. & Traugott, E. C. (1993). *Grammaticalization.* Cambridge: Cambridge University Press.

Katz, Y. & Shifman, L. (2017). Making Sense? The Structure and Meanings of Digital Memetic Nonsense. *Information Communication and Society* 20 (6): 825–842. DOI: 10.1080/1369118X.2017.1291702

Kecskes, I. & Zhang, F. (2009). Activating, Seeking and Creating Common Ground: A Socio-Cognitive Approach. *Pragmatics & Cognition* 17, 2, 331–355.

Kövecses, Z. (1990). *Emotion Concepts.* New York: Springer.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal About the Mind.* Chicago, London: University of Chicago Press.

Leeson, L. and Saeed, J. I. (2012). *Irish Sign Language. A Cognitive Linguistic Account.* Edinburgh: Edinburgh University Press.

Linell, P. (1998). *Approaching Dialogue: Talk, Interaction, and Contexts in Dialogical Perspective.* Amsterdam: John Benjamins.

Milner, R. M. (2012). *The World Made Meme: Discourse and Identity in Paricipatory Media.* PhD Dissertation, University of Kansas.

Miltner, K. (2014). 'There's No Place for Lulz on Lolcats': The Role of Genre, Gender and Group Identity in the Interpretation and Enjoyment of an Internet Meme. *First Monday*, 19(8). doi:10.5210/fm. v19i8.5391.

Nissenbaum, A. & Shifman, L. (2015). Internet Memes as Contested Cultural Capital: The Case of 4chan's \B\ Board. *new media & society*, 1-19. doi: 10.1177/1461444815609313.

Ruiz de Mendoza Ibáñez, F. J. & Galera Masegosa, A. (2014). *Cognitive Modeling. A Linguistic Perspective.* Amsterdam: Benjamins.

Shifman, L. (2013). Memes in a digital world: Reconciling with a conceptual troublemaker. *Journal of Computer-Mediated Communication,* 18 (3), 362-377.

Ungerer, F. & Schmid, H-J. (1996). *An Introduction to Cognitive Linguistics.* London / New York: Longman.

Wittgenstein, L. (1960). Philosophische Untersuchungen. In: Wittgenstein, L. *Tractatus logico-philosophicus (=Schriften 1),* Frankfurt am Main: Suhrkamp.

LingBaW *Linguistics Beyond And Within 4 (2018), 38-50*

Pre-D non-possessor positions in Hungarian*

Judit Farkas and Krisztina Karácsonyi
University of Pécs, Hungary

Abstract

The paper investigates pre-D non-possessor positions in Hungarian. In Hungarian, non-deverbal nominal constructions containing pre-D non-possessor positions are acceptable only if they contain a demonstrative pronoun and also an adjective, and the appearance of a pre-D possessor does not impact the acceptability of the sentence. The paper also gives a brief discussion of similar constructions with pre-D non-possessors in German, mainly to shed light on the Hungarian data. Although German also allows for pre-D non-possessors, it does so under different conditions. A short topicalized element can readily appear in German sentences as a non-possessor dependent, but in this language a possessor can never appear in the same noun phrase. The paper also discusses deverbal nominal constructions with pre-D non-possessor dependents in Hungarian. In these constructions the presence of a possessor argument is indispensable. This is due to the fact that the placement of the non-possessor argument in a position preceding the possessor is legitimized by the fact that the former takes scope over the latter within the internal information structure of the matrix noun phrase. The paper also deals with the syntactic structure of said deverbal nominals.

Keywords: pre-D non-possessors, Hungarian, German, deverbal nominals

1. Introduction

This paper investigates pre-D non-possessor positions in Hungarian, first in the light of findings currently proposed for German. In both languages it is possible for a non-possessor dependent to appear before the definite article. The conditions, however, are different across the two languages.

The phenomenon is all the more interesting considering the general agreement amongst Hungarian mainstream linguists that nothing can appear before the pre-D possessor position, not at least as an element forming one constituent with the noun head (see Bartos (2000), for instance). However, as is shown in (1), this is not the case.

* Supported by the ÚNKP-17-2-I.-PTE-284 New National Excellence Program of the Ministry of Human Capacities, project EFOP-3.4.3-16-2016-00005 and NKFIH 128518 (The Syntax of Hybrid Categories).

(1) a. (?)Na például [*mindkét városba* ugyanannak a futárnak az elküldése],
well for_instance both city.Ill same.Dat the courier.Dat the sending.Poss
az hiba volt.
that mistake was
'Well for instance, as for sending the same courier *to both cities*, that was a mistake.'

b. (?)Na például [*Olaszországból* az a finom bor],
well for_instance Italy.Ela that the delicious wine
az nagyon drága volt.[1]
that very expensive was
'Well for instance, as for that delicious wine *from Italy*, that was very expensive.'

In the Hungarian examples we use a constituency test (offered by Alberti, Farkas, & Szabó (2015)) based on the for instance-construction. This contrastive topic construction[2] can be completed with a resumptive pronoun (such as az 'that'),[3] which signals the end of the nominal constituent tested, while the sequence na például 'well for instance' signals the beginning of the tested construction.

In section 1, German constructions containing pre-D non-possessor dependents will be analyzed, followed by the investigation of similar Hungarian constructions. Section 2 deals with Hungarian deverbal nominal constructions, specifically, factors conditioning the occurrence of pre-D non-possessor dependents. In section 3, a syntactic structure is offered for the constructions discussed in section 2. The paper concludes with a short summary (section 4).

2. Pre-D non-possessors in German and in Hungarian

Pre-D non-possessors do not exclusively appear in Hungarian, they can also be found in German; see Roehrs's (2014) examples in (2):

---

[1] The following abbreviations are used in the glosses:
(i) case suffixes: acc(usative), dat(ive), ela(tive), gen(itive), ill(ative), ine(ssive), sub(lative);
(ii) gender suffixes: msc (masculine), fem(inine), neu(tral);
(iii) other suffixes on nouns: poss (possessedness suffix), poss.1sg/poss.2sg (possessive agreement suffixes);
(iv) affixes on verbs: 1sg (agreement suffix), perf (a perfectivizing verbal prefix (preverb));
(v) derivational suffixes: adj(evtivalizer), inf(initive), part(iciple);

Throughout the whole paper, the following six-degree scale of grammaticality judgments, given in Broekhuis, Keizer, & Dikken (2012, viii), is used: *: unacceptable; $^{*?}$: relatively acceptable compared to *; $^{??}$: intermediate or unclear status; $^{?}$: marked: not completely unacceptable or disfavored; $^{(?)}$: slightly marked, but probably acceptable. We also follow Broekhuis, Keizer, & Dikken (2012, xiv) in using introspective judgments of the two authors (both native speakers of Hungarian) as the criterion of what word orders are part of the language associated with what readings (cf. Featherston, 2007, section 5.4).

[2] The contrastive topic status of the construction in question can be verified by the fact that it is "pronounced with a particular, contrastive intonation comprised of a brief fall and a long rise" (É. Kiss, 2002, p. 22).

[3] Note that in Hungarian the form of the resumptive pronoun *az* 'that' is identical with that of the definite article (which can also appear as *a* if the next word in the noun phrase begins with a consonant). Note also that the noun phrase with a pre-D non-possessor can also contain the demonstrative pronoun *az* 'that' (as in (1b), for instance), the form of which is identical with the resumptive pronoun. The former is the part of the noun phrase under investigation but the latter is not. For the sake of clarity, the borders of the noun phrase construction with a pre-D non-possessor are marked by square brackets.

(2) *a. aus Italien* der Wein
Ela Italy the.Msc wine
'the wine *from Italy*'
a'. *aus Spanien* das Obst
Ela Spain the.Neu fruit
'the fruit *from Spain*'
b. *nach Hamburg* der Zug
after Hamburg the.Msc train
'the train *to Hamburg*'
b'. *nach Ostern* die Woche
after Easter the.Fem week
'the week *after Easter*'

As is shown in the series of examples in (3), there are similar (non-deverbal) constructions in Hungarian as well. The conditions, however, are slightly different in the two languages. In German, short topicalized dependents can readily appear in this position followed by a definite article and a noun head. In Hungarian, the constructions with the same structure are unacceptable; see examples (3a) and (3b).

(3) a. *?Na például [*Olaszországból* a bor], az nagyon drága volt.
well for_instance Italy.Ela the wine that very expensive was
'Well for instance, as for the wine *from Italy*, that was very expensive.'
b. *?Na például [*Hamburgba* a vonat], az nagyon drága volt.
well for_instance Hamburg.Ill the train that very expensive was
'Well for instance, as for the train *to Hamburg*, that was very expensive.'

A German-like construction, however, is almost fully acceptable in Hungarian if it contains a demonstrative pronoun and an attributive adjective, as is shown in examples (4a-b).

(4) a. (?)Na például [*Olaszországból* az a finom bor], az nagyon drága volt.
well for_instance Italy.Ela that the delicious wine that very expensive was
'Well for instance, as for that delicious wine *from Italy*, that was very expensive.'
b. ?Na például [*Hamburgba* az az új expresszvonat], az nagyon drága volt.
well for_instance Hamburg.Ill that the new express_train that very expensive was
'Well for instance, as for that new express train *to Hamburg*, that was very expensive.'

The construction is fully acceptable in Hungarian if the pre-D non-possessor is not a simple proper name but a noun phrase with a more complex structure containing a demonstrative pronoun and attributes (5a-b). The reason for this may lie with the higher degree of specificity[4] of the matrix noun phrase in the latter case, which can be regarded as a legitimizing factor to fill in the position in question (cf. Alberti, 1997; Farkas, & Alberti, 2018, pp. 11–13).

(5) a. Na például [*abból a bájos olasz kisvárosból* az a finom bor],
well for_instance that.Ela the lovely Italian small_town.Ela that the delicious wine

[4] Specificity here is understood as the type of discourse prominence in which the addressee is (also) assumed to know the given entity referred to. The phenomenon in question, however, requires future research, in order to set it in Givón's (1983) specificity hierarchy and in Heusinger's (2011) more complex system.

az nagyon drága volt.
that very expensive was
'Well for instance, as for that delicious wine *from that lovely small town in Italy*, that was very expensive.'

b. Na például [*abba a messzi német kisvárosba* az az új expresszvonat],
well for_instance that.Ill the far German small_town.Ill that the new express_train
az nagyon drága volt.
that very expensive was
'Well for instance, as for that new express train *to that German small town far from here*, that was very expensive.'

It is worth investigating if and how the appearance of a possessor impacts the acceptability of these constructions. In German, a possessor cannot appear in such constructions either in a pre-D position (6a), or postnominally (6b).

(6) a. *aus Italien *Peters* Wein
from Italy Peter.Gen wine
Intended meaning: '*Peter's* wine from Italy'

b. ??aus Italien der Wein *von Peter*
from Italy the.Msc wine of Peter
Intended meaning: 'the wine *of Peter* from Italy'

In contrast to German, the appearance of a pre-D possessor has no impact on the acceptability of the construction in Hungarian, as is shown in examples (7a-b) compared to examples (4a) and (4b).

(7) a. (?)Na például [Olaszországból *Luiginak* az a finom bora],
well for_instance Italy.Ela Luigi.Dat that the delicious wine.Poss
az nagyon drága volt.
that very expensive was
'Well for instance, as for that delicious wine *of Luigi* from Italy, that was very expensive.'

b. ?Na például [Hamburgba *a MÁV-nak* az az új expresszvonata],
well for_instance Hamburg.Ill the MÁV-Dat that the new express_train.Poss
az nagyon drága volt.
that very expensive was
'Well for instance, as for that new express train *of the Hungarian State Railways* to Hamburg, that was very expensive.'

To sum up the requirements, a short topicalized element can appear as a non-possessor dependent in German while a possessor cannot appear in the construction. In Hungarian, however, the construction in question is acceptable only if it contains a demonstrative pronoun and an attributive adjective, while the appearance of a pre-D possessor has no impact on the acceptability of the construction.

3. Hungarian deverbal nominal constructions

This section discusses deverbal nominal constructions in Hungarian. A pre-D non-possessor dependent can appear in Hungarian in certain deverbal nominal constructions (8a). As is

shown in (8b), however, a simple topicalized proper noun cannot appear here (cf. the German examples in (2)).

(8) a. Na például [*mindkét konferenciára* ugyanannak a kollégának
well for_instance both conference.Sub same.Dat the colleague.Dat
az elküldése], az hiba volt.
the away_sending.Poss that mistake was
'Well for instance, as for sending the same colleague *to both conferences*, that was a mistake.'

b. *Na például [*Pécsre Péternek az elküldése*], az hiba volt.
well for_instance Pécs.Sub Péter.Dat the away_sending.Poss that mistake was
Intended meaning: 'Well for instance, as for sending Péter *to Pécs*, that was a mistake.'

We propose that the placement of the non-possessor argument in a position preceding the possessor is legitimized by the former taking scope over the latter within the internal information structure of the matrix noun phrase, and, at least within the prenominal zone, there is no other possibility to express this scope by word order. In what follows, the principal grounds for this assumption will be provided.

In Hungarian, word order always corresponds with the scope hierarchy of the arguments, as is shown by É. Kiss (2002, pp. 113–115). What can be observed in examples (9a-b) is that if we reverse the order of the two quantifiers, the meaning changes parallel to the change in scope hierarchy.

(9) a. [Minden kollégámat]QUANTIFIER [többször is]QUANTIFIER meghívtam.
every colleague.Poss.1Sg.Acc several_times also invited.1Sg
'It holds for each of my colleagues that I invited him/her several times.'

b. [Többször is]QUANTIFIER [minden kollégámat]QUANTIFIER meghívtam.
several_times also every colleague.Poss.1Sg.Acc invited.1Sg
'I invited all of my colleagues (together) several times.'

É. Kiss (2002, p. 47) refers to this phenomenon as Scope Principle, "according to which operators have scope over the domain they c-command" (É. Kiss 2002, p. 47)) and in Hungarian this scope order corresponds with the surface word order. Karácsonyi (2017) pointed out that this Scope Principle is not only valid for finite verbs but is a general pre-head principle. As is demonstrated in the series of examples in (10), the principle also holds for infinitives and participles. If the order of the quantifiers in these constructions is reversed, the meaning also changes parallel to the change in scope hierarchy, just as in the case of finite verbs.

(10) a. Na például [[minden kollégát] [többször is] meghívni],
well for_instance every colleague.Acc several_times also invite.Inf
az túlzás volt.
that exaggeration was
'Well for instance, as for inviting each colleague several times, that was an exaggeration.'

a'. Na például [[többször is] [minden kollégát] meghívni],
well for_instance several_times also every colleague.Acc invite.Inf
az túlzás volt.
that exaggeration was
'Well for instance, as for inviting all colleagues several times, that was an exaggeration.'

b. Na például [a [minden kollégát] [többször is] meghívó] Ili, ő fura.
well for_instance the every colleague.Acc several_times also invite.Part Ili she strange
'Well for instance, as for Ili, who invited each colleague several times, she is strange.'

b'. Na például [a [többször is] [minden kollégát] meghívó] Ili, ő fura.
well for_instance the several_times also every colleague.Acc invite.Part Ili she strange
'Well for instance, as for Ili, who invited all colleagues several times, she is strange.'

Next, let us consider the question of scope hierarchy in the case of noun phrases. In Hungarian, both the dative case-marked possessor and the unmarked possessor precede the prenominal position capable of hosting non-possessor dependents (11a).

(11) a. Na például [[mindkét kollégá(nak] a) [többszöri] meghívása],
well for_instance both colleague.Dat the several_times.Adj invitation.Poss
az túlzás volt.
that exaggeration was
'Well for instance, as for inviting both colleagues several times (not necessarily together), that was an exaggeration.'

b. Na például [a [többszöri] meghívása [mindkét kollégának]],
well for_instance the several_times.Adj invitation.Poss both colleague.Dat
az túlzás volt.
that exaggeration was
Meaning 1: 'Well for instance, as for inviting both colleagues several times (not necessarily together), that was an exaggeration.'
Meaning 2: 'Well for instance, as for inviting both colleagues (together) several times, that was an exaggeration.'

In a case like this, the following question arises: How can the reverse scope order be obtained. The first option is to place the possessor dependent after the noun head, as in (11b). In this case, however, the resulting construction is ambiguous: it can be associated both with the meaning where the quantified possessor takes scope over the quantified non-possessor dependent (see meaning1) and with the meaning where the non-possessor takes scope over the possessor (see meaning2).

The ambiguity of the constructions containing post-head operators can also be observed in the case of finite verbs (12), as is shown by É. Kiss (1992, pp. 163–164).

(12) Ili hívta meg [minden kollégáját] [többször is].
Ili invited perf every colleague.Poss.Acc several_times also
Meaning 1: 'It was Ili who invited each of her colleagues several times (not necessarily together).'
Meaning 2: 'It was Ili who invited all of her colleagues (together) several times.'

In the case of noun phrases, it is exactly the pre-D non-possessor position that guarantees that the Scope Principle prevail. Example (13a) is not ambiguous: it can only be associated with the meaning reflecting the scope hierarchy.

(13) a. Na például [[többször is] [mindkét kollégának] a meghívása],
well for_instance several_times also both colleague.Dat the invitation.Poss
az túlzás volt.
that exaggeration was
'Well for instance, as for inviting both colleagues (together) several times, that was an exaggeration.'

b. *Na például [[többször is] a meghívásod], az túlzás volt.
well for_instance several_times also the invitation.Poss.2Sg that exaggeration was
Intended meaning: 'Well for instance, as for inviting you several times, that was an exaggeration.'
b'. *Na például [[többször is] a meghívás], az túlzás volt.
well for_instance several_times also the invitation that exaggeration was

As is shown in (13b-b'), the presence of an explicit possessor is indispensable in such noun phrases: the construction containing a pre-D non-possessor dependent is unacceptable if it "only" contains a possessor which is *pro*-dropped (13b) or if it contains no possessor (13b'). The explicitness of the possessor, while obligatory, is not the exclusive condition. As is shown in (14), only a possessor with a relevant-set based operator function can appear in such constructions, that is, noun phrases with focus or quantifier functions (see also (8b)). As was mentioned in connection with example (8), the same holds for non-possessor dependents.

(14) a. *Na például [[többször is] [Péternek] a meghívása],
well for_instance several_times also Péter.Dat the invitation.Poss
az túlzás volt.
that exaggeration was
Intended meaning: 'Well for instance, as for inviting Péter several times, that was an exaggeration.'

As for the number of pre-D non-possessor dependents, Karácsonyi (2017) points out that what hosts such dependents is not a single pre-D position but a zone which can host more than one non-possessor dependent (15).

(15) a. Na például [[mindkét évben] [mindkét konferenciára]
well for_instance both year.Ine both conference.Sub
ugyanannak a kollégának az elküldése], az hiba volt.
same.Dat the colleague.Dat the away_sending.Poss that mistake was
'Well for instance, as for sending the same colleague to both conferences in both years, that was a mistake.'
b. Na például [[mindkét évben] [mindkét témában] [mindkét konferenciára]
well for_instance both year.Ine both topic.Ine both conference.Sub
ugyanannak a kollégának az elküldése], az hiba volt.
same.Dat the colleague.Dat the away_sending.Poss that mistake was
'Well for instance, as for sending the same colleague to both conferences in both topics in both years, that was a mistake.'

To sum up the requirements concerning the pre-D non-possessor position in the case of deverbal nominals in Hungarian: a position in question is acceptable only when its filler serves as a quantifier or focus and is followed by a(n overt) possessor also functioning as a quantifier or focus. This is due to the fact that the pre-D non-possessor position can only be used if a non-possessor dependent takes scope over the possessor (which otherwise precedes the non-possessor element(s)). Filling in the pre-D non-possessor position is the only way to unambiguously express the [non-possessor > possessor] scope order.[5] This is in total harmony

[5] One of the two anonymous reviewers raised a hypothesis according to which operator movement in Hungarian simply serves to disambiguate scope hierarchy. As example (i) shows, this is not the case, since movement in

with the characteristic feature of Hungarian according to which word order always reflects scope order.

4. The syntactic structure of deverbal nominals containing a pre-D non-possessor dependent in Hungarian

This section discusses the structure of Hungarian deverbal nominals containing a pre-D non-possessor dependent. Figure 1 below represents the structure of (the relevant part of) the noun phrase in (13a), repeated here as (16).

(16) *többször* *is* mindkét kollégának a meghívása
several_times also both colleague.Dat the invitation.Poss
'inviting both colleagues (together) *several times*'

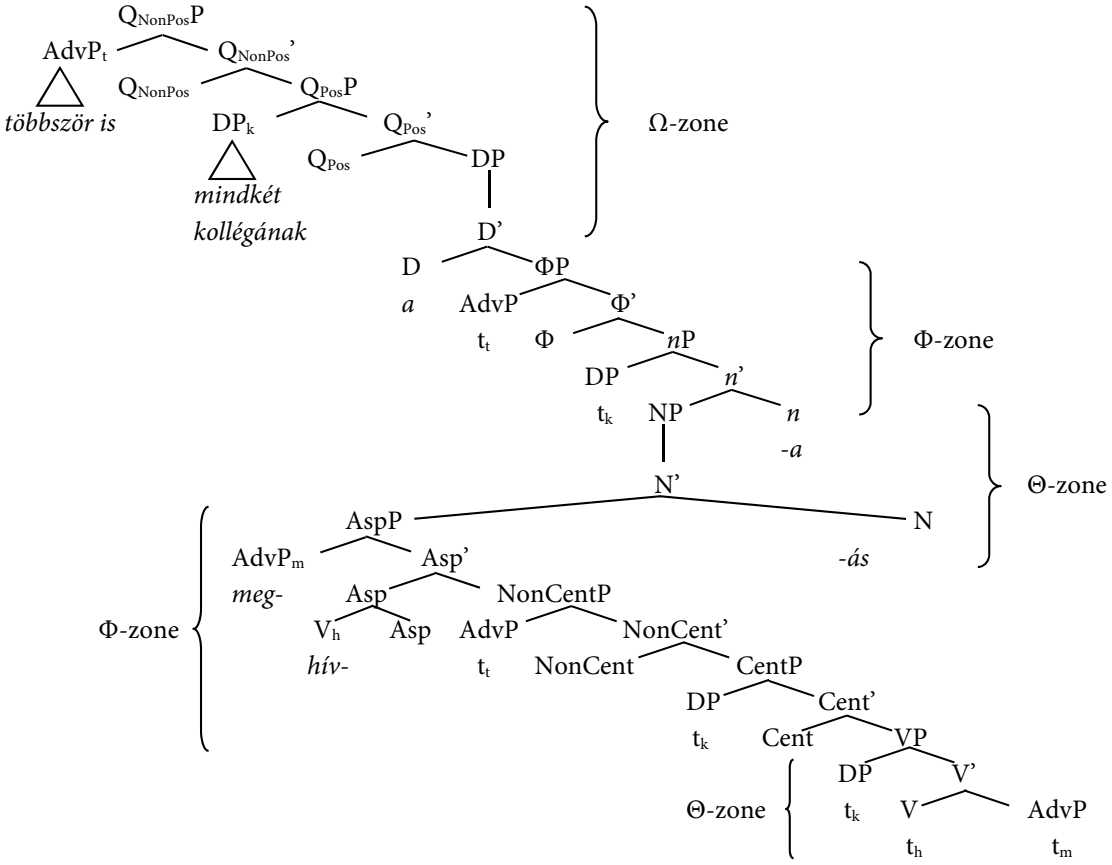


Figure 1: *The syntactic structure of the noun phrase in (16)*

In order to capture the phenomenon of internal-scope taking within nominal expressions, we propose a general syntactic representation in which the essentially morphology-based Hungarian traditions are integrated with Giusti's (1996) cartographic Split-DP Hypothesis (see

Hungarian also happens when there is only one operator. Thus, movement is required for operators to take their scope.

(i) [Mindkét kollégát] elbocsátotta a főnök.
both colleague.Acc send_away.Past.3Sg the boss
'Both colleagues were fired by the boss.'

also Farkas, Szabó, & Alberti, 2017). The result is a tripartite nominal structure consisting of thematic domains ($\Theta_V\Delta$, $\Theta_N\Delta$), agreement domains ($\Phi_V\Delta$, $\Phi_N\Delta$) and discourse domains ($\Omega_V\Delta$, $\Omega_N\Delta$) according to Grohmann's (2003, p. 211 (37b)) theory of Prolific Domains.

We constructed the layer hierarchy of the structure of the nominal expression as is shown in Figure 1 essentially on the basis of the proposals by Bartos (2000) and É. Kiss (2002, pp. 151–180). However, it required immense notational and some derivational modification to adapt it to Grohmann's (2003, pp. 227–228) two basic generalizations over movement or derivational dependencies within tripartite clause-like cycles: (i) cycle-internal movement always targets the next higher domain (according to this order: $\Omega\Delta \leftarrow \Phi\Delta \leftarrow \Theta\Delta$), and (ii) movement across cycles targets a position within the same type of domain in the next higher cycle (i.e., $\Omega\Delta \leftarrow \Omega\Delta$, $\Phi\Delta \leftarrow \Phi\Delta$, $\Theta\Delta \leftarrow \Theta\Delta$). Thus we accept Grohmann's Anti-Locality Hypothesis (2003, p. 26), according to which "movement must not be too local".

Bartos (2000, pp. 678–683), by referring to Baker's (1985) Mirror Principle, proposed layers between D and N essentially on the basis of the assumption that morphology is frozen syntax, that is, "today's morphology is yesterday's syntax" (Givón, 1971, p. 413). The morphology of the noun head in (16) suggests that (i) a PossP-layer builds upon the NP-layer, reflecting the change resulting in a "possessed noun", and (ii) then comes a NumP-layer for numeral information, and (iii) then an Agr_NP-layer, given the agreement between possessor and possessee in number and person. In Figure 1, (i) PossP is referred to as *n*P on the analogy between the (non-thematic) argument generating function of Poss/*n* and the Agent "generating" function of *v*,[6] (ii) there is no NumP-layer as the singular number does not require its creation, (iii) there is no agreement layer because this kind of agreement is asymmetrical, or defective, in the sense that in possessive structures with non-pronominal possessors there is no agreement (Bartos, 2000, pp. 678–683).

Our tree building method is completed with the consistent consideration of a principle of D-visibility.[7] The principle declares that either the specifier or the head of the DP must be spelled out. In Figure 1, the default filler of the DP appears, namely the definite article *a(z)* 'the' in the head of D.

The example in (16) shows the following characteristic property of complex-event denoting deverbal nominal constructions: they can have internal information structure. In order to capture this special capability, we need an extended DP structure which integrates the morphological (Mirror-Principle-based (Baker, 1985)) Hungarian traditions (Szabolcsi & Laczkó, 1992; Bartos, 2000; É. Kiss, 2002) with the cartographic Split-DP Hypothesis (Giusti, 1996; Ihsane, & Puskás, 2001) by assuming noun-phrase-internal operator layers (see Grohmann, 2003, p. 211 (37b), Farkas, & Alberti, 2016, and Alberti, Farkas, & Szabó, 2017).

As the given noun phrase is a highly verbal nominal expression, namely a complex-event denoting deverbal nominal construction, an appropriately extended VP-structure based upon

---

[6] The use of *n*P in Hungarian was proposed by Giuliana Giusti (p.c., 25 May 2016).

[7] Its application to Hungarian on the basis of a proposal by Alexiadou (2004, p. 47) is convincingly argued for by Egedi (2015, p. 6), among others. Something similar, namely that economy forces in some languages to have a zero D when (Spec,DP) is occupied by an overt element and to have a filled D when (Spec,DP) is non-overt or not filled at all, is proposed in different works by Giusti on Romanian (e.g., Giusti, 2005, p. 37) as an Economy Principle.

an embedded verb *hív* 'invite' as its head is assumed to be taken by the nominalizer *-Ás* in the noun head. The idea that a deverbal nominalizer occupies the position of the N head in the center of the deverbal nominal construction and takes a projection containing a VP is essentially based on Fu, Roeper, & Borer (2001). We consider this embedded verbal construction in the depth of the noun phrase to be the "scope-semantic" source of the internal scope. The reason why it is possible in Hungarian for an internal information structure to be hosted in noun phrases is that the Hungarian noun-phrase structure is (even) more flexible than was hypothesized earlier.

The preverb is base-generated in the complement position of the verb, from which position it moves to the specifier of AspP in the **Φ**-domain as the element responsible for the given (perfective) aspect. The head of the AspP is occupied by the verb stem.[8]

In the particular nominal expression in Figure 1, the embedded verb has one argument (besides the preverb), which is base-generated in the specifier of the embedded VP. From this position, it is raised into the specifier of the (Spec,CentP) as a central argument of the verb. Then, from this **Φ**-position, it moves to the nominal hemisphere, where it appears as a possessor. The movement from a **Φ**-position (belonging to the verb) to another **Φ**-position in the nominal domain is in harmony with Grohmann's (2003, pp. 227–228) generalization over movement across cycles (the moving element targets a position within the same type of domain in the next higher cycle). In the case of nominals with the deverbal nominalizer *-ás*, prototypically the Theme argument is designated for this role. The possessor is thus first raised into (Spec,*n*P), which layer is responsible for checking (the mere fact of) possessedness. In harmony with this, the *n* head is occupied by the possessedness suffix *-a*. Due to the suffix status of the morpheme, this phrase is right-headed just as the N head occupied by the deverbal nominalizer. After the Theme argument moves to the specifier of the little *n*P, which belongs to the **Φ**-zone of the noun, it rises further to a pre-D layer, since it is a NAK possessor. The target position is in the **Ω**-domain of the nominal hemisphere, namely the specifier of the lower quantifier, the $Q_{Pos}P$,[9] due to the operator character of the phrase in question.[10]

The last phrase to be discussed is the free adjunct *többször is* lit. 'also several times'. Being a non-central dependent of the verb, this phrase is base-generated in the specifier of a NonCentP as a free adjunct—in total harmony with Grohmann's (2003, p. 313 (O1b)) theory. From this **Φ**-position it moves to another **Φ**-position in the nominal domain (named **Φ**P in Figure 1).[11] As the free adjunct in question is also a quantifier, it rises to an **Ω**-position in the

[8] There is no projection containing T(ense)P (see Alberti, 2004; É. Kiss, 2006; 2008) because deverbal nominal constructions obligatorily contain even exclusively-perfectivizing preverbs (see Laczkó, 2000, pp. 314–316) but they express no tense.

[9] In earlier models (Szabolcsi & Laczkó, 1992; Bartos, 2000), nak-possessors rise to (Spec,DP). This position, however, is not suitable for a model such as this one; we need a finer-grained layer structure if we also intend to satisfy the principle of D-visibility (according to which the head and the specifier of the DP cannot be filled in within the same structure).

[10] According to our analysis, beside interclausal movements, clause-internal ones can also take place in constructions like (16). Intervention effects in Hungarian DPs have not become apparent yet.

[11] The idea of **Φ**P is based on Ihsane & Puskás (2001, p. 45), whose approach is based on Aboh's (1998) ideas, in assuming that (potentially iterable) functional projections can be inserted between the DP-layer and the NP/*n*P-layer in the Hungarian DP-structure.

nominal hemisphere, namely to the specifier of the upper quantifier, that is, the specifier of the $Q_{NonPos}P$. Indices 'Pos' and 'NonPos' are needed to ensure the order of the operator projections, because, as is shown in (17a), the non-possessor dependent must precede the possessor in the pre-D zone, while the reversed order is unacceptable.[12] The reason for this may lie with the fact that reversed scope order can be expressed by using an attributive construction (17b) and there is no reason to use a marked structure.

(17) a. *mindkét kollégának *többször* *is* a meghívása
both colleague.Dat several_times also the invitation.Poss
Intended meaning: 'inviting both colleagues (not necessarily together) *several times*'

b. mindkét kollégának a *többszöri* meghívása
both colleague.Dat the several_times.Adj invitation.Poss
'inviting both colleagues (not necessarily together) *several times*'

5. Summary

The paper discussed pre-D non-possessor positions in Hungarian. These positions can be found in two types of constructions in the language.

On the one hand, a pre-D non-possessor can appear in non-deverbal nominals in the case of which the construction is acceptable only if it contains a demonstrative pronoun and an attribute, and the appearance of a pre-D possessor has no impact on the acceptability of the construction. The acceptability of this type of construction increases in correlation with the higher degree of specificity of the matrix noun phrase, which can be regarded as a legitimizing factor to fill in the position in question. It was briefly discussed in connection with this type that similar constructions can also be found in German. Nevertheless, the conditions are different in the two languages. In German, it is possible for a short topicalized element to appear as a non-possessor dependent, while a possessor cannot appear in the construction.

On the other hand, pre-D non-possessor dependents can appear in certain deverbal nominal constructions in Hungarian. In this case the presence of a possessor argument is indispensable, and the placement of the non-possessor argument in a position preceding the possessor is legitimized by scope visibility, prevalent in Hungarian: it is only in this way that the non-possessor argument can unambiguously take scope over the possessor within the internal information structure of the matrix noun phrase. The paper also proposed a syntactic structure for complex-event denoting deverbal nominals containing pre-D non-possessor dependents, which is in total harmony with Grohmann's (2003) theory of Anti-Locality.

As for future research, it would be worthwhile to investigate other Germanic languages or different language families to see if they have pre-D non-possessors. In case of a positive answer, it should be explored whether the conditions in those languages are similar to the ones that can be detected in German or in Hungarian, or if they are different from those discussed in this paper.

[12] Farkas, Szabó & Alberti (2017) offers a more abstract, that is, less Hungarian-NP-specific, solution to the operator order observed in the pre-D zone, which, however, requires a radically new (but promising) approach to phonological Transfer.

References

Aboh, E. (1998). *On the Syntax of Gungbe Noun Phrases*. Clearinghouse on Languages and Linguistics ELIC Documentation Reproduction service No ED 420 209.

Alberti, G. (1997). Restrictions on the Degree of Referentiality of Arguments in Hungarian Sentences. *Acta Linguistica Hungarica,* 44 (3-4), 341–362.

Alberti, G. (2004). Climbing for Aspect – with no Rucksack. In K. É. Kiss & H. van Riemsdijk (Eds.), *Verb Clusters; A study of Hungarian, German and Dutch.* Linguistics Today 69 (pp. 253–289). Amsterdam, The Neatherlands / Philadelphia, USA: John Benjamins.

Alberti, G., Farkas, J., & Szabó, V. (2015). Argument for Arguments in the Complement Zone of the Hungarian Nominal Head. In K. É. Kiss, B. Surányi, & É. Dékány (Eds.), *Approaches to Hungarian 14* (pp. 3–36). Amsterdam, The Neatherlands: John Benjamins.

Alberti, G., Farkas, J., & Szabó, V. (2017). Internal-scope taking arguments in the information structure of deverbal nominals in Hungarian. In H. van der Hulst & A. Lipták (Eds.), *Approaches to Hungarian 15* (pp. 1–33). Amsterdam, The Neatherlands: John Benjamins.

Alexiadou, A. (2004). On the development of possessive determiners. Consequenses for DP structure. In E. Fuss & C. Trips (Eds.), *Diachronic clues to synchronic grammar*. Linguistik Aktuell 72 (pp. 31–58). Amsterdam, The Neatherlands: John Benjamins.

Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry, 16*, 373–415.

Bartos, H. (2000). Az inflexiós jelenségek szintaktikai háttere [The syntactic background of inflectional phenomena]. In F. Kiefer (Ed.), *Strukturális magyar nyelvtan 3* [A Structural Grammar of Hungarian 3] (pp. 653–762). Budapest, Hungary: Akadémiai Kiadó.

Broekhuis, H., Keizer, E., & Dikken, M. den (2012). *Syntax of Dutch – Nouns and Noun Phrases*. Amsterdam, The Neatherlands: Amsterdam University Press.

É. Kiss, K. (1992). Az egyszerű mondat szerkezete [The structure of the simple clause]. In F. Kiefer (Ed.), *Strukturális magyar nyelvtan* 1 [A Structural Grammar of Hungarian 1] (pp. 79–177). Budapest, Hungary: Akadémiai Kiadó.

É. Kiss, K. (2002). *The Syntax of Hungarian.* Cambridge, UK: Cambridge University Press.

É. Kiss, K. (2006). The function and the syntax of the verbal particle. In K. É. Kiss (Ed.), *Event Structure and the Left Periphery* (pp. 17–56). Dordrecht, The Neatherlands: Springer.

É. Kiss, K. (2008). Free word order, (non-)configurationality, and Phases. *Linguistic Inquiry, 39*, 441–475.

Egedi, B. (2015). The Hungarian definite article across time: an account of its spreading. Talk at International Conference on the Structure of Hungarian 12, Leiden, 22-23 May 2015. (Handout: http://www.nytud.hu/oszt/elmnyelv/fonevikif.html.)

Farkas, J., & Alberti, G. (2016). Positions for oblique case-marked arguments in Hungarian noun phrases. *Jezikoslovlje, 17* (1-2), 295–319.

Farkas, J., & Alberti, G. (2018). Characterization. In G. Alberti & T. Laczkó (eds.), *Nouns and noun phrases* (pp. 5–151). Amsterdam, The Neatherlands: Amsterdam University Press.

Farkas, J., Szabó, V., & Alberti, G. (2017). Information-Structurally (Un)Ambiguous Nominal Constructions in Hungarian. *Finno-Ugric Languages and Linguistics 6* (2), 38–80.

Featherston, S. (2007). Data in generative grammar: The stick and the carrot. *Theoretical Linguistics 33* (3), 269–318.

Fu, J., Roeper, T., & Borer, H. (2001). The VP within Process Nominals: Evidence from Adverbs and the VP Anaphor *Do-So*. *Natural Language & Linguistic Theory 19*, 549–582.

Giusti, G. (1996). Is there a FocusP and a TopicP in the Noun Phrase structure? *Working Papers in Linguistics, 6* (2), 105–128.

Giusti, G. (2005). At the Left Periphery of the Romanian Noun Phrase. In M. Coene & L. Tasmowski (eds.), *On Space and Time in Language* (pp. 23–49). Romania, Cluj Napoca: Clusium.

Givón, T. (1971). Historical syntax and synchronic morphology: an archaelogist's field trip. *Chicago Linguististics Society, 7,* 394–415.

Givón, T. (1983). Topic Continuity in Discourse. An Introduction. In T. Givón (ed.), *Topic Continuity in Discourse. A Quantitative CrossLanguage Study* (pp. 1–41). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Grohmann, K. K. (2003). *Prolific Domains: On the Anti-Locality of Movement Dependencies.* Linguistik Aktuell 66. Amsterdam, The Neatherlands: John Benjamins.

Heusinger, K. von. (2011). Specificity, Referentiality and Discourse Prominence: German Indefinite Demonstratives. In I. Reich et al. (eds.), *Proceedings of Sinn & Bedeutung 15* (pp. 9–30). Saarbrücken, Germany: Universaar – Saarland University Press.

Ihsane, T., & Puskás, G. (2001). Specific is not Definite. *Generative Grammar in Geneva, 2,* 39–54.

Karácsonyi, K. (2017). A pre-D nem birtokosi pozíció [The pre-D non-possessor position]. In G. Böhm, D. Czeferner, & T. Fedeles (Eds.), *Specimina Operum Iuvenum 5* (pp. 59–87). Pécs, Hungary: PTE.

Laczkó, T. (2000). Az ige argumentumszerkezetét megőrző főnévképzés [Argument structure preserving noun derivation]. In F. Kiefer (Ed.), *Strukturális magyar nyelvtan 3* [A Structural Grammar of Hungarian 3] (pp. 293–407). Budapest, Hungary: Akadémiai Kiadó.

Roehrs, D. (2014). Splitting the Nominal Left Periphery into DP and TopP: Two Pre-adjectival Positions for Possessives in German, http://ling.auf.net/lingbuzz/002111. downloaded: 05.10.2015)

Szabolcsi, A., & Laczkó, T. (1992). A főnévi csoport szerkezete [The Structure of the Noun Phrase]. In F. Kiefer (Ed.), *Strukturális magyar nyelvtan* 1 [A Structural Grammar of Hungarian 1] (pp. 179–298). Budapest, Hungary: Akadémiai Kiadó.

LingBaW *Linguistics Beyond And Within 4 (2018), 51-62*

The use of abbreviations by superscript letter in an early fifteenth-century manuscript of the Wycliffite Bible*

Joanna Grzybowska
John Paul II Catholic University of Lublin, Poland

Abstract

The use of scribal abbreviations in medieval manuscripts was mainly dictated by the need to save space and time as the creation of a medieval book was both extremely costly and time-consuming. One of the types of scribal abbreviations used in medieval manuscripts is abbreviation by superscript letter. In this type of abbreviations one superscript letter indicates the ending of a given word, or, in some cases, a medial position. Both vowels and consonants were used as abbreviations by superscript. They usually denoted, apart from the actual letter written in superscript, the preceding vowel or the letter <r>. According to Cappelli (1929/1982), superscript letters in Latin were used mainly in word-final positions; however, it was not uncommon for a superscript vowel to appear word-medially.

The main objective of this paper is to investigate the use of superscript letters in an early fifteenth-century manuscript of the Wycliffite Bible (Mscr.Dresd.Od.83) on the basis of the Gospel of Matthew. Within the manuscript there are both superscript consonants and vowels. However, in some cases these abbreviations seem to appear in very specific contexts, whereas in other cases the contexts allowing the abbreviations to appear are much broader. The possible reasons behind this situation will be discussed within this paper along with the correspondence between the superscript letter and the spelling conventions used within the manuscript.

Keywords: superscript letter, Wycliffite Bible, scribal abbreviation, medieval manuscript

1. Introduction

The main objective of this paper is to investigate the use of abbreviations by superscript letter in the Dresd.Od.83 manuscript of the Wycliffite Bible, the problems connected to the classification of these abbreviations, and spelling issues connected to expanding them.

One of the most striking phenomena that characterise medieval manuscripts were scribal abbreviations. They can be treated as scribe's intervention to the original text as they can be

* I would like to thank Professor Magdalena Charzyńska-Wójcik for all her comments on this paper.

often expanded in more than one way, which, in turn, may lead to possible misinterpretations. The main purpose for using abbreviations was the need to save time and space, as the process of creating a medieval *codex* was time-consuming and required the use of expensive materials. The number of scribal abbreviations increased with time and, by the Late Middle Ages, created a complex system. Scribes abbreviated both whole words and letters.

1.1. The system of scribal abbreviations in Late Middle Ages – An overview

The practice of using abbreviations by scribes is older than Christianity and can be traced back to the Roman times, when, for instance, legal formulae as well as first names were abbreviated in inscriptions (Bischoff, 1986/2012, p. 150). The fall of the Western Roman Empire contributed to the disappearance of the universal style of writing and, consequently, to the disappearance of the common way of abbreviating letters in manuscripts, which led to regional divisions and, therefore, confusion.

In the Early Middle Ages, the abbreviations used by scribes differed depending on the region of Europe the scribes came from. Another turning point in the history of scribal abbreviations came with the reign of Charlemagne, who introduced several reforms into the writing system (Bischoff, 1986/2012, p. 114). The so-called Carolingian reform worked as a unifying factor in the system of scribal abbreviations (Bischoff, 1986/2012, p. 154). It seems that Charlemagne was especially concerned about the state of biblical manuscripts, as he did not allow young men who were training to be a scribes to copy the Scripture (Kenyon, 1895/1898, p. 182). What is more, he extended an invitation to Alcuin, a monk from England, who was asked to create a corrected version of Vulgate (Charzyńska-Wójcik, 2013, p. 14; Kenyon, 1895/1898, p. 182). It is important to mention that the Carolingian 'renovatio' concerned many different areas of learning. Within the biblical manuscripts Alcuin was mainly concerned with correcting the orthography as well as the organisation of the text on the manuscript page (McKitterick, 2005, p. 155). The scribes started using the standardised writing style (later called the Carolingian minuscule) and abbreviation system (Bischoff, 1986/2012, p. 154). When it comes to the system of abbreviations, the regional differences were reduced and the amount of scribal abbreviations started to correlate to the status of the copied text – the higher the status, the smaller amount of abbreviations was used (Bischoff, 1986/2012, p. 154).

The system of scribal abbreviations seemed to be quite stabilised by the end of the eleventh century, even though scribes still added new abbreviations to it. The number of scribal abbreviations used in manuscripts grew steadily throughout the medieval period to become fully developed by the High Middle Ages. The Late Middle Ages can be seen mainly as a period of consolidating and refining the system.

Bischoff (1986/2012, p. 156) argues that because of the clarity of the system at the end of the medieval period, it is possible to divide the abbreviations into four groups: abbreviations using conventional signs, suspensions, contractions and syllabic abbreviations. At the same time, a different way of classifying scribal abbreviations is proposed by Cappelli (1929/1982, p.1), who claims that they can be divided into: "truncations, contractions, abbreviation marks significant in themselves, abbreviation marks significant in context, superscript letters and

conventional signs". This difference in classification may stem from the fact that, even though the system of abbreviations seemed quite stabilised and uniform by the end of the medieval period, it still varied between scribes or scriptoria. However, the term "syllabic abbreviations" seems to be quite unclear and misleading because other groups of abbreviations mentioned by Bischoff (1986/2012) can be used, in some contexts, to abbreviate syllables. For example, the so called 9-mark (℘), which is classified as a conventional sign, abbreviates sequences of letters <con> and <cum>. Therefore, for the sake of this study, it seems best to use the division proposed by Cappelli (1929/1982).

Truncations, according to Cappelli (1929/1982, pp. 1-7), are types of abbreviations in which only the first part of the word is spelled out and the abbreviation mark replaces the final letters. The missing part of the word can be abbreviated either by general signs (e.g. macrons) or specific truncation signs, which state precisely what kind of letters are abbreviated.

Contractions, on the other hand, are the type of abbreviations in which an abbreviation mark, usually a macron, stands for one or more of the middle letters, which is why abbreviations by contraction can be classified as abbreviations of whole words. Paoli (1891; as cited in Cappelli, 1929/1982, p. 7) claims that abbreviations by contraction can be divided into pure contractions, in which only first and last letters are spelled out, and mixed contractions, in which also middle letters are spelled out. Among abbreviations by contraction are both *nomina sacra* and abbreviations of hierarchical terms (e.g. 'episcopus').

Abbreviation marks significant in context are a type of abbreviation which represents different letters depending on the letters which appear closest to it (Cappelli, 1929/ 1982, p. 18). For instance, when the so-called 3-mark ʒ follows the letter <o>, it usually stands for <-nem>, but it can also stand for the letter <m>. However, in Roman cursive from the twelfth century the 3-mark usually stands for the letters <-rum> and is written at the same level with other letters within the line (Cappelli, 1929/1982, p. 22). Interestingly, this abbreviation mark may correspond to 'macron', which also stands for the letter <m>. This shows that two different abbreviation marks sometimes could represent the same letter. To slightly complicate matters, the opposite situation was also possible.

Among the abbreviations described by Cappelli (1929/1982) there are abbreviation marks significant in themselves. It is a type of abbreviation mark which does not change its meaning depending on the context it appears in (Cappelli, 1929/1982, p. 13). A good example of this type of abbreviation is a macron, which was used both as an abbreviation for the letters <m> and <n> and an abbreviation of two adjoining syllables with the letters <-i-> and <-o-> in words such as 'fi*nis*', 'ho*min*is' or 'ho*no*rem' (Bischoff, 1986/2012, p. 158).

The next group of abbreviations used by medieval scribes are represented by superscript letter — a type of abbreviation in which a letter written above a consonant indicates the ending of a given word. Scribes used both vowels and consonants as superscript letter. Superscript vowels can stand for both the vowel and for that vowel and the letter <r> (Cappelli, 1929/1982, p. 30). For instance, superscript <a> and <i> could stand respectively for <-ra> and <-ri> as illustrated in: 'sup^{a}' for 'sup*ra*' or 'p^{i}or' for 'p*ri*or' (Bischoff, 1986/2012, p. 159). Superscript consonants were generally used in a similar way to vowels and abbreviated both the consonant and the preceding vowel.

Abbreviations using conventional signs include abbreviation marks which are not recognizable as letters and, according to Cappelli (1929/1982), are used as isolated signs used instead of a well-known word or phrase. As examples of this type of abbreviations Cappelli (19929/1982, p. 39) shows signs *9* and *ↄ,* which stand for the Latin words (and prefixes) *con* and *cum*.

1.2. Manuscripts of the Wycliffite Bible

The term 'Wycliffite Bible' is, in fact, a term which can be assigned to two translations of the Bible. The first version of Wycliffe's Bible, usually referred to as the Early Version, was created in the year 1380 and it was the translation of the New Testament. Two years later, in 1382, a translation of the whole Bible was completed (Kenyon, 1895/1898, p. 200). The second version, often called the Late Version, was most likely created around 1388. This revision of the first translation was not as literal as the Early Version. It is argued by Norton (2000, p. 7) that the translators "moved towards a more readable English rendering, one more obviously capable of standing by itself without reference back to the Latin".

When it comes to the manuscripts of the Wycliffite Bible, it is important to mention that quite a large number of them survived, even though the copies of Wycliffe's Bible were systematically destroyed due to being claimed as heretical. The notion that the Wycliffite Bible was associated with heresy is connected to the fact that Wycliffe openly expressed his reformatory views on doctrinal matters which were not supported by the Pope or church dignitaries. In effect, Wycliffe's works were often not in accordance with the teaching of the Church. In fact, Wycliffe's first quarrel with ecclesiastical authorities was connected with the Pope's demands for tribute money, which were refused by the Parliament in 1366 (Kenyon, 1895/1898, p. 199). In the year 1378 Wycliffe wrote the treatise *De Potestate Papae*, which pointed to the Scripture as the sole source of doctrine and, therefore, it undermined the authority of the Pope (Metzger, 2001, p. 56). What is more, *De Civili Dominio,* another of Wycliffe's works, resulted in Wycliffe facing charges of heresy and, in the end, condemnation by the Pope (Ghosh, 2004, p. 22) As is stated by Estep (1986, p. 66), one of Wycliffe's claims, condemned by Gregory XI, was the denial of transubstantiation, one of the key dogmas of the Church. Wycliffe claimed that transubstantiation was impossible from both logical and metaphysical point of view (Olson, 1999, p. 358).

Wycliffe's opposition towards some of the dogmas of the Church resulted in the ban on the Wycliffite Bible issued 1408 (Deanesly, 1920, p. 131). However, it should be mentioned that translating the Bible into a vernacular language was not generally prohibited, because translations into French were common in the pre-Wycliffite times (Deanesly, 1920, p. 206). The official condemnation of the translation did not seem to influence its status among the readers. Interestingly, Poleg (2013) argues that it is possible to find formal similarities between many manuscripts of the Wycliffite Bible and those of the Latin Vulgate —similarities which point to the need for an English translation within the orthodox community. At the same time, "[t]he removal of the General Prologue and of marginal glosses dissociated Wycliffite Bibles from the realm of Lollard thought, while simultaneously supporting an unmediated access to the biblical

text" (Poleg, 2013, pp. 90-91). It meant that the Wycliffite Bible was read not only by Lollards, but also by other people, who wanted to gain better understanding of the Bible.

2. Analysis of superscript letters in the Dresd.Od.83 manuscript of the Wycliffite Bible

2.1. Methodology

The analysis of abbreviations by superscript letter is based on my own transcript of the Gospel of Matthew from the Dresd.Od.83 manuscript of Wycliffe's Bible. The transcript consists of 24 620 words (100 060 characters, excluding spaces). The transcript was prepared with the view to representing the manuscript as accurately as possible with respect to all scribal conventions. Obviously all abbreviations were expanded but the expanded sequences were subsequently italicised[1]. Another convention used while preparing the transcript concerns the use of punctuation. The system of punctuation marks used in the transcript reflects the one used by the scribe. Another editorial decision made while preparing the transcript of the Gospel concerned the letters used in the manuscript. In the manuscript there are some letters, namely 'thorn' and 'yogh', which are remnants of the Old English alphabet. Both of these letters are used in transcript, however, it is important to mention that the 'thorn' used in the manuscript does not have an ascender (which it used to have in the Anglo-Saxon times).[2]

After preparing the transcript, all abbreviations were extracted and grouped according to the letters they represented. Afterwards, the occurrences for each group were counted. All contexts which excluded abbreviations were also catalogued just as those in which the use of an abbreviation was permitted by the context but the scribe wrote the word in full.

2.2. Description of the Dresd.Od.83 manuscript of the Wycliffite Bible

The Dresd.Od.83 manuscript of Wycliffe's Bible is dated back to 1400[3]. Currently it is held in the Sächsische Landesbibliothek – Staats – und Universitätsbibliothek Dresden (SLUB Dresden). The manuscript contains selected books of the Bible: the Gospels, the Epistles, the Book of Revelation as well as Jerome's prologues to these books. What is more, the manuscript contains the Calendar of Gospel Readings and the Readings from the Old Testament.

The manuscript is written on parchment. Its measurements[4] are 17 x 12 cm. The first page of each book is illuminated with flowers, whereas other pages are not. It is important to mention that flower motifs in illumination are often used in the manuscripts of that century (Fisher, 2004, p. 5). When it comes to the first pages of individual chapters, the decoration is limited to red lines starting from the blue initial.

---

1 Because all abbreviations are marked in italics, all lexical items within this article are given in single quotation marks

2 In effect, the thorn used in the analysed manuscript is similar to another letter used in Old English texts, namely 'wynn' <ƿ>. It is important to realise, however, that the similarity is purely visual and does not imply an affinity of any other type.

3 According to the information provided by SLUB Dresden.

4 According to the information provided by the SLUB Dresden.

The text is written in two columns with initials and pilcrows written in blue ink. The 'capitula' are also written in blue ink and they are followed by red abbreviations of the evangelist's name. The manuscript is written in gothic textura, a type of script used by scribes in the High and Late Middle Ages, which could be characterised by angular shapes of letters (Bischoff, 1986/2012, pp. 133-134). Another feature of the script used in the manuscript is the joining of letters, especially <m>, <n> and <u>, which consist of either two or three minims. The scribe differentiates between lower and upper case letters. However, it is important to mention that capitalisations are used inconsequently.

Scribal mistakes can be observed throughout the manuscript; however, most of these errors were corrected by the original scribe (the script used by the scribe who corrected the mistakes is exactly the same as the one used by the original scribe). It is necessary to mention that scribal mistakes are intrinsic features of manuscripts, which were caused by the way medieval manuscripts were produced.

2.3. Abbreviations by superscript letter in the Dresd.Od.83 manuscript of Wycliffe's Bible

Cappelli (1929/1982, p. 30) claims that abbreviation by superscript letter is an abbreviation in which the superscript letter stands for the missing letters at the end of the word. The superscript letters can be both vowels and consonants. Within the Gospel of Matthew of the Dresd.Od.83 manuscript of the Wycliffite Bible the scribe used four different superscript letters. They indicated seven different sequences of letters. What is more, the superscript letters used within the manuscript can be divided into certain groups.

2.3.1. Superscript <t>

Superscript <t> can be expanded into four different sequences of letters: <at>, <ut>, <ot> and <ip>. Examples of use of those abbreviations are shown in (1) below.

(1) a. verse 5.39 "But I seie to ȝou. <u>pa*t*</u>[5] ȝe aȝenstonde not a*n* yuel man"
b. verse 3.15 "b<u>u*t*</u> *iesu* answeride *and* seide to hi*m*/ Suffre now/ For pus it fallip to us to fulfille al riȝtfulnes/"
c. verse 9.36 And he siȝ pe peple. *and* hadde rupe on hem; for pei weren trauelid; *and* liggynge as scheep <u>*not*</u> hauynge a scheepherde/"
d. verse 13.17 "p*at* pe profecie of isaie seiyn=ge. be fulfillid in he*m*; <u>*wip*</u> heeringe ȝe schule*n* heere; *and* ȝe shulen not vndir=stoned/ and ȝe seynge schulen se; *and* ȝe shulen not se/"

It is important to mention that a superscript <t> may appear in all positions within the line of a manuscript; it is not restricted to the margin. At the same time, it can also be noticed that it is used only in the word-final position. The Figure 1 below shows the number of occurrences of each sequence of letters abbreviated by superscript letter <t>.

[5] The underlined words illustrate the use of superscript letters within the context of a verse.

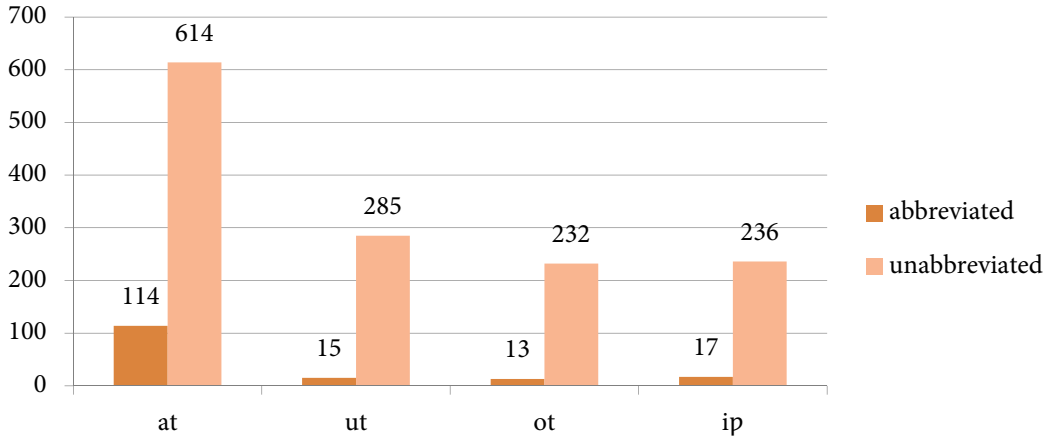


Figure 1: *Abbreviation for letters <at>/<ut>/<ot>/<ip>*

It is important to mention that superscript <t> does not appear in equal numbers in all contexts. Another important fact connected to this superscript letter is that it appears in strictly specified contexts, to the effect that it is limited to specified words. For instance, when it represents the letters <at>, the abbreviation appears only in the word 'pat'. The abbreviation for the letters <ut> appears in the word 'but', for <ot> – 'not' and for <ip> – in the word 'wip'. The sequence of letters <ip> stands out from the remaining sequences abbreviated by this sign. In this case, the sequence of letters abbreviated by the letter <t> does not contain this letter in it and deserves a more detailed description.

As mentioned in Section 2.1., the scribe used the letter 'thorn' <p> in the manuscript. In fact, there seems to be a preference for this letter in comparison with <th>. Within the Gospel of Matthew there are 3783 occurrences of <p> and only 79 occurrences of <th>. The examples in (3) and (4) present some of the contexts in which <th> and <p> appear.

(3) a. verse 1.3 "Iudas bigat Fares *and* zaram of Tha=mar/ Fares bigat esrom/"
b. verse 1.13 "Zorobabel bi=gat Abiuth/ Abiuth bi=gat Eliachym/ Eliachi*m* bigat Asor/"
c. verse 2.6 "and pei seiden to him/ In beth=leem of Iuda/ for so it is writen bi aprofete"

(4) a. verse 1.1 "The book of pe ge=neracioun of *iesu* crist pe sone of dauip. pe sone of Abra=ham/"[6]
b. verse 2.4 "But king eroude her=de/ *and* was troublid; *and* al *ierusalem* wip hi*m*/"
c. verse 4.3 "*and* pe temptere cam nyȝ: *and* seide to hi*m*/ If pou art g=oddis sone; seie pat pese stoones be maad looues/"

It can be deduced from the examples provided in (3) and (4) that the <th> sequence seems to be used more often with proper names, whereas <p> is used more often in other contexts. This is supported by the numbers presented in Figure 2 below. It is additionally interesting because the digraph <th> seems to appear mostly in proper names, which simultaneously constitute words of foreign origin. The digraph appears 46 times in proper names (such as 'Obeth', 'Joathan', 'bethleem' or 'sathanas'), and 33 times in other words (mostly in the word 'the' – 19 times). It is important to mention that eight instances of <th> appear in words beginning with a capital letter both in proper names, for example 'Thamar', and in words like 'the' or 'therfore',

[6] The <th> sequence in this verse is used probably due to the illumination of letter <t> as it is the first verse of the Gospel of Matthew.

when they appear at the beginning of a verse or chapter. It seems significant especially in comparison with the letter <p>, which appears only once within a proper name – 'Dauip'.

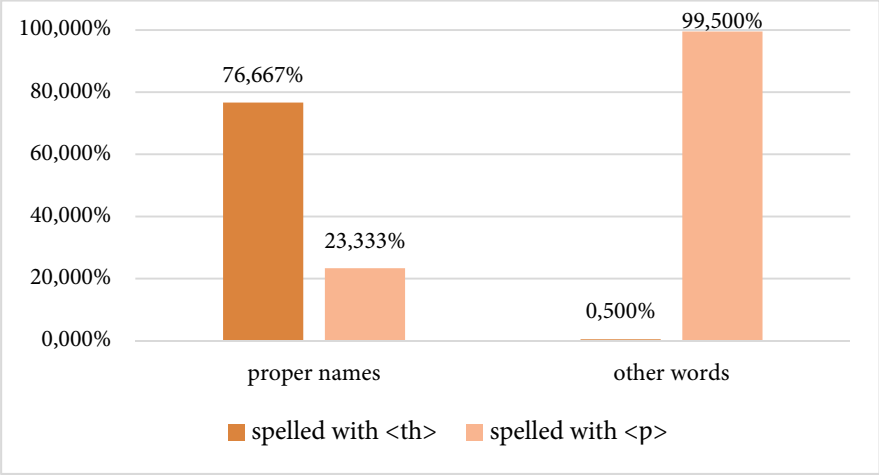


Figure 2: *Percentage of Proper Names and Other Words Spelled with Letters <th> and <p>*

Therefore, the fact that superscript <t> was used by the scribe to abbreviate the ending of the word 'wip' seems surprising. The scribe's choice to use superscript <t> in this context, instead of a different type of abbreviation does not correspond to the spelling of the word 'wip' within the manuscript. In fact, there is only one instance of 'with' spelled with a digraph in the manuscript – verse 20.20: "¶ panne pe modir of pe sones of ze=bedee came to him with hir sones honourynge. *and* axynge sum pi*n*g of hi*m*/". In comparison, the unabbreviated word 'wip', spelled with the thorn, appears 122 times within the Gospel of Matthew.

Therefore, a question arises about the relationship between <th> and <p> in the manuscript, especially, because there seems to be no clear reason for using superscript <t> as an abbreviation for the statistically preferred letter <p>. However, it should be mentioned that <p> is a runic character and, because of that, does not belong to the same writing system as the purely Latin scribal abbreviations. Therefore, it is possible that the scribe chose letter <t> because the context excluded <p>. In fact, there are several instances of digraph <th> appearing in contexts excluding the appearance of the thorn (for example, initials, beginnings of chapters).

2.3.2. Superscript <s>

The next abbreviation of this type is superscript <s>, which stands for the sequence of letters <is>. The examples shown in (5) represent all types of contexts, in which this abbreviation appears.

(5) a. verse 6.2 "nyle p*ou* trumpe bifore pee as y=pocritis doon i*n* synagog*is*"
b. verse 22.4 "Eftsoo=nes he sente op*er*e seruau*n*=tis; *and* seide/ Seie ȝe to pe men pat ben beden to pe feeste/ lo I haue maad re=di my mete; my bolis *and* my volatilis ben slay*n*. *and* alle pingis ben redy; come ȝe to pe weddingis/"
c. verse 24.3 "And whanne he satt on pe hil of olyuete; hi=se disciplis camen to hi*m* p*r*iuyli *and* seiden/ Seie to us whanne pese ping*is* schulen be; *and* what toke=ne of pi comynge. *and* of pe eendinge of pe world/"

d. verse 24.30 "*and* panne pe toke=ne of mannes sone sc=hal appere in heuene; *and* panne alle kynred*is* of pe erpe schulen wei=le *and* pei schulen se man=nes sone comynge in pe cloudis of heuene; wip miche ve*r*tu *and* maies=te/"
e. verse 26.60 "and pei fou*n*den not; whanne manye fals witness*is* weren comen/ But at pe laste two false wit=nessis camen;"
f. verse 6.12 "*and* forȝeue to vs oure dettis. as *and* we forȝeuen to oure dettour*is*/"
g. verse 7.24 "perfore ech ma*n* p*at* heerip pese my word*is and* doip hem; schal be m*aa*d lijk to awiseman p*at* hap bildid his hous on astoo*n*."
h. verse 9.4 "and whanne *iesu* hadde seen her pouȝt*is*. he seide/ wherto penken ȝe yuel pingis in ȝoure hertis.?"
i. verse 13.16 "But ȝoure iȝen pat seen ben blessid; *and* ȝoure eer*is* pat heeren/"
j. verse 15.15 "Petir answeride; *and* seide to hi*m*/ Expowne to us p*is par*able/"
k. verse 25.19 "For of pe herte goon out yuele thouȝtis. mansleyngis. auoutries. fornicaciouns. peftis. fals witnessing*is*. blasfemies/"
l. verse 21.41: "And pei seien to hi*m*/ He schal lese yuel. pe yuel men/ *and* he schal sette to hire his vy=neȝerd to op*e*re erthetilier*is*; whiche schulen ȝelde fruitis to him in her tymes/"

In the Gospel of Matthew superscript <s> appears fifteen times in twelve different words. This differentiates it from superscript <t>, which was lexically restricted depending on the sequence of letters it stood for. However, there is also another difference between this abbreviation and the previous one. Whereas superscript <t> may appear in any position within the line of the manuscript, superscript <s> seems to be used only at the end of a line. Therefore, it appears that the scribe chose to use the abbreviation only in places where using the unabbreviated sequence of letters would disrupt the composition of the page. All 15 occurrences of superscript <s> appear near the margin, where the use of the unabbreviated form would result in lengthening the line and, in that way, disrupting the way the page looks.

2.3.3. Superscript <u>

The third superscript vowel used within the manuscript is <u>, which stands for sequence of letters <ou>. The examples in (6) show the contexts in which the abbreviation appears within the manuscript.

(6) a. verse 19.19 "He seip to hi*m*; whiche/ And *iesus* seide/ p*ou* schalt not do mansleyin=ge; pou schalt not do a=uoutrie. pou schalt not do pefte pou schalt nott seie fals witnessyynge."
b. verse 5.32 "But I seie to ȝ*ou*. pat eueri man pat leeuep his wijf. outtaken cause of fornicacioun; makip hir to do letcherie/ *and* he pat weddip peforsaken wijf; doip auoutrie/"
c. verse 14.21 "And pe noumbre of me*n* pat eeten was fyue p*ou*=sand of men; outaken wy*m*men *and* litil children/"

The abbreviation appears 30 times within the Gospel of Matthew out of 219 occurrences of this sequence of letters. Interestingly, it is the only superscript letter which appears word-medially. However, it should be mentioned that the word 'pousand' which the superscript <u> appears in was divided by the scribe in order to preserve the layout of the page.

2.3.4. Superscript <e> and <o>

The next abbreviation by superscript letter is superscript <e>, which stands for the sequence of letters <ee>. The examples shown in (7) present the use of this abbreviation in the Dresd.Od.83 manuscript.

(7) a. verse 6.4 "p*at* pin almes be in hid=lis *and* pi fad*ir* p*at* seep i*n* hid=lis. schal quite *pee*/"
b. verse 8.13 "And *iesus* seide to pe cen=turioun/ go; *and* as p*ou* hast bileeued. be it doon to p*ee*/ *and* pe child was heelid fro pat hour"
c. verse 9.22 "and *iesu* turned and siȝ hir *and* seide/ Douȝtir haue p*ou* trist; pi feip hap maad p*ee* saaf/ and pe wo*m*ma*n* wa*s* hool fro pat hour/"
d. verse 26.11 so fadir; for so it was plesinge tofore p*ee*/

There are 15 occurrences of that abbreviation; all of them appear in the word 'pee'. It is especially interesting, because there seem to be other words, in which the abbreviation could appear in, for example 'ȝee', 'tree', 'galilee', or 'iudee' yet it is not attested there.

The fact that superscript <e> appears only within one word poses the question whether this sign abbreviates a sequence of letters, as the previously mentioned superscript <t> and <s> do, or whether the p^{e} sign should be treated as an abbreviation of a whole word. However, it is possible to explain why superscript <e> does not appear in the context of other words without assuming it is an abbreviation for the whole word. One explanation might be that the scribe chose to use a different abbreviation to express <ee>. For example, in other words ending with the letters <ee> there is a letter 'r' preceding them, which means the whole sequence of letters <ree> can be abbreviated by the mark ˀ written above the preceding letter. The only exceptions seem to be the words like 'ȝee', 'citee', 'galilee' and 'iudee'. However, when it comes to proper names like 'galilee' or 'iudee' the scribe usually chooses not to abbreviate them or use an abbreviation for the whole word – a contraction or an abbreviation mark significant in itself (Section 1.1.). Although, the fact that the scribe chose to use this abbreviation in a single word could suggest classifying it as abbreviation for the whole word rather than for a sequence of letters, the clarity of the system should be taken into consideration — a matter I will return to in the concluding section.

The superscript letter with the smallest number of occurrences is <o>. It represents the sequence of letters <ro>. The examples in (8) show the use of this abbreviation within the manuscript.

(8) a. verse 4.21 "*and* he ȝede forp f*ro* pat place; *and* he siȝ tweȝe opere brip*er*en. James of ze=bede *and* Ioon his brop*ir*. in a schip wip zebede her fad*ir* amendinge her nettis/ *and* he clepide hem/"
b. verse 24.29 "¶ And anoon aftir pe tribulacioun of po dai=es; pe su*n*ne schal be m*aa*d derk; *and* pe moone schal not ȝeue hir liȝt; *and* pe sterris schulen falle f*ro* heuene. *and* pe vertues of heuenes schule*n* be mo=ued/"

Interestingly, superscript <o> is used only twice as an abbreviation for the ending of the word 'fro'. This shows a huge disproportion in comparison with the 99 occurrences of unabbreviated word 'fro'. It seems, however, that the abbreviation is restricted to the position at the end of a line (before the margin) as it never appears in the middle of the line. Another reason behind the fact that the scribe did not use the abbreviation in any other context might be that there is no

other word within the Gospel of Matthew which ends with the sequence of letters <ro>. Moreover, a very similar sign is used by the scribe to indicate the beginning (number) of a new chapter (c^0) — e.g. "¶xviii^{o}". Even though this abbreviation occurs only at the beginning of a chapter and is written in red ink, the graphic similarities between it and superscript <o> could potentially cause some misunderstandings.

3. Conclusions

To sum up, the abbreviations by superscript letter used in the Dresd.Od.83 manuscript of the Wycliffite Bible are not a homogenous group. They can be divided into three categories depending on how they are used within the manuscript.

The first type of abbreviation, superscript <t>, represents the largest amount of the sequences of letters: <at>, <ut>, <ot> and <ip>. They appear in very restricted contexts; each of them is used only in one specific word, for example 'but' or 'pat'. The second type, superscript <s>, differs quite significantly from superscript <t>. Although it represents only one sequence — <is>, it is used in many words. The first superscript vowel discussed in this paper — <u> stannds for one sequence of letters — <ou>. It appears in three contexts, in words 'ȝou', ''pou' and 'pousand'. It is the only superscript letter which appears word-medially; however, it should be noticed that even in this context, the abbreviation appears near the margin, as the word 'pousand' was divided by the scribe. The next group of superscript letters contains superscripts <e> and <o>. They represent only one sequence of letters each and appear in strictly specified contexts. Even though, at first, it may seem that these abbreviations were used differently than the two other superscript letters, they should not be categorised as abbreviations of whole words. This is mainly connected to the fact that, within the Gospel of Matthew, there are either no other words which would provide the suitable context for the appearance of a superscript letter (for example, sequence <ro> abbreviated by superscript <o>) or there are other abbreviations which may be used in place of the superscript (for instance, sequence of letters <ree> or words such as 'galilee'). Throughout the Gospel of Matthew, a clear division can be observed between abbreviations which can be classified as abbreviations of whole words and those which represent abbreviations of letters or sequences of letters. All other abbreviations of whole words within the manuscript are either contractions (e.g. 'nomina sacra') or "abbreviation marks significant in themselves" (for example &). What is more, it should be mentioned that neither of the abbreviations of whole words were used as abbreviations of sequences of letters. Because of that classifying all abbreviations by superscript letter as abbreviations of sequences of letters instead of classifying some of them as abbreviations of whole words seems more rational from the point of view of the whole system of abbreviations in the Gospel of Matthew, especially when there are other reasons for not using superscript letters in more contexts, such as a possibility of using different abbreviations or lack of words ending with a certain sequence of letters

Source

The Wycliffite Bible = Das neue Testament englisch. (1400). Dresden, Sächsische Landesbibliothek -Staats- und Universitätsbibliothek. Mscr.Dresd.Od.83.

References

Bischoff, B. (2012). *Latin Palaeography: Antiquity and the Middle Ages*. (D. Ó Cróinín & D. Ganz, Trans.). New York, NY: Cambridge University Press. (Original work published1986).

Cappelli, A. (1982). *The Elements of Abbreviation in Medieval Latin Paleography*. (D. Heimann & R. Kay, Trans.). Lawrence, KS: University of Kansas Libraries. (Original work published 1929).

Charzyńska-Wójcik, M. (2013). *Text and Context in Jerome's Psalter: Prose Translations into Old, Middle and Early Modern English*. Lublin: Wydawnictwo KUL.

Deanesly, M. (1920). *The Lollard Bible and Other Medieval Biblical Versions*. Cambridge:University Press.

Estep, W. R. (1986). *Rennaissance and Reformation*. Grand Rapids, MI.: William B. Eerdmans Publishing Company.

Fisher, C. (2004). *Flowers in Medieval Manuscripts*. Toronto: Toronto University Press.

Ghosh, K. (2004). *The Wycliffite Heresy: Authority and the Interpretation of Texts*. Cambridge: Cambridge University Press.

Kenyon, F.G. (1898). *Our Bible and the Ancient Manuscripts*. (3rd ed.). London: Eyre and Spottiswode. (Original work published 1895).

McKitterick, R. (2005). The Carolingian Renaissance of Culture and Learning. In J. Story (Ed.), *Charlemagne: Empire and Society* (pp. 151-166). Manchester & New York, NY: Manchester University Press.

Metzger, B. M. (2001). *The Bible in Translation. Ancient and English Versions*. GrandRapids, MI: Baker Academic.

Norton, D. (2000). *A History of the English Bible as Literature*. Cambridge: Cambridge University Press.

Olson, R. E. (1999). *The Story of Christian Theology: Twenty Centuries of Tradition &Reform*. Downers Grove, IL: InterVarsity Press.

Paoli, C. (1891). *Le abbreviature nella paleografia latina del medio evo: saggio metodico-pratico*. Florence: Sansoni.

Poleg, E. (2013). Wycliffite Bibles as Orthodoxy. In S. Corbellini (Ed.), *Instructing the Soul, Feeding the Spirit and Awakening the Passion. Cultures of Religious Reading in the Late Middle Ages (*71-91*). T*urnhout: Brepols.

LingBaW *Linguistics Beyond And Within 4 (2018), 63-78*

Auxiliary clitics in Polish*

Dorota Jagódzka
Palacký University Olomouc, Czech Republic

Abstract

Polish auxiliary clitics constitute an interesting set of data which draws attention to cross-linguistic differences among Slavic languages. A general principle for clitic placement in Indo-European languages is the one described by Jacob Wackernagel in his 1892 work. He concluded that clitics appeared in the second position in the clause, after the first word in a sentence. This pattern was true to some degree in Old Church Slavonic[1] and still holds for a number of contemporary Slavic languages e.g. Serbo-Croatian, Slovene, Czech and Slovak which have second position clitics. Bulgarian and Macedonian have verb adjacent pronominal clitics and Polish has auxiliary clitics (Migdalski 2007, 2010, Pancheva 2005). Also in the older versions of Polish language the above mentioned tendency was strong.

In Modern Polish auxiliary clitics attach to the *l*-participle most frequently. However, one of the unusual properties they possess is the ability to choose almost every clausal element for their host. Polish auxiliary clitics can trigger morphophonological alternations on their hosts, which is an affix-like property; however, at the same time they display clearly clitic-like behaviour when they attach freely to words of any lexical class.

The aim of this paper is to present and analyze the morpho-syntactic properties of two kinds of auxiliary clitics: bound and free. The bound clitics carry person-number agreement markers for past tense (the so called 'floating' or 'mobile' inflections). The free clitic is the morpheme *by* used for conditional and subjunctive mood.

Keywords: clitics, Slavic languages, inflection

1. Introduction

One of the first attempts to classify clitic systems cross-linguistically was Wackernagel's (1982) study on Indo-European word-order. He analysed the behaviour of unstressed elements in Ancient Greek, Sanskrit Latin and Germanic and concluded that they show a strong preference

* This research was funded by Ministry of Education, Youth and Sports of the Czech Republic (MŠMT in Czech), grant IGA_FF_2017_003 (Mezinárodní kontext pro angloamerickou literaturu, lingvistiku a translatologii) and IGA_2018_012 (Internacionalizace v oblasti angloamerické literární vědy, lingvistiky a translatologie).
I would like to thank two anonymous reviewers for comments and suggestions. All the remaining mistakes are mine.

[1] Migdalski (2007, 2010) reports that the Wackernagel type of clitics did not have an entirely uniform distribution in OCS, and that the clearer case of second position clitics in this language was the operator clitic.

for second position in the clause, after the first constituent. Second position, however, may mean different things in different languages – this is either following the first word in a sentence (Tagalog), the first word or the first constituent of any sort (Serbo-Croatian) or the first constituent as long as it is immediately dominated by S (Walbiri).

Slavic languages constitute a rich field for research as they exhibit different cliticization patterns. The majority of South and West Slavic languages have second position cliticization. Polish has weak pronouns and auxiliary clitics. Bulgarian and Macedonian, similarly to Germanic and Romance languages, have verb-adjacent clitics which means that the verb is always a host for a clitic (Migdalski 2007, 2010, 2016, Pancheva 2005, Van Riemsdijk 1999).

Another typology created from a cross-linguistic perspective based on data from many unrelated languages comes from Zwicky (1977) and analyses clitics as phonologically-weak elements which differ in syntax and in the relationship to their full forms. He introduced a two-way distinction between simple and special clitics. Simple clitics have regular distribution (identical to their free, stressed counterparts), are phonologically-weak and must adjoin to another word in order to get the prosodic support. Special clitics exhibit a more idiosyncratic distribution.

Migdalski (2016) came out with yet another classification. He divided clitics into generalized clitics and operator clitics. Generalized second position clitics comprise auxiliary clitics and pronominals. Operator clitics are a special group of Wackernagel clitics. They form a distinct class because all of them share one property, namely expressing the illocutionary force of a clause. Operator clitics are attested in all Slavic languages, irrespective of whether they have other second position clitics, verb-adjacent clitics, weak pronouns or no other clitics at all. Some of the examples of operator clitics are: Czech *prý*, which is used to report non-witnessed events; interrogative complementizer *li* (present in many contemporary Slavic languages such as Bulgarian, Macedonian, Russian, Serbo-Croatian); and operator clitic *že* in Czech and Russian, and *że* in Polish which mark the focus of a sentence.

1.1. Polish auxiliary clitics

Polish has an interesting stet of auxiliary clitics which do not exhibit the Wackernagel's second position behaviour (Migdalski 2016). They can attach to various elements in the clause and do not require second position distribution (for more detailed description of their distribution see section 3.1 in this paper). For those reasons they are sometimes called in the related literature 'floating inflections' (Spencer 2000). Polish bound auxiliary clitics carry person and number features, and when affixed to the *l*-participle, they agree with the subject of the clause. This is why they are also called Person-Number Agreement (Embick 1995) or simply 'clitic auxiliaries' (Borsley and Rivero 1994). To illustrate the pattern of how the auxiliary clitics attach to the verb, I present Table 1 below which shows the past tense forms of the verb 'to be' in Polish together with the appropriate endings:

***Table 1**: Past tense of verb być – 'to be'.*

SG			
Person	MSC	FEM	NT
1. 'I was'	*by-ł-e-m*	*by-ł-a-m*	-
2. 'You were'	*by-ł-e-ś*	*by-ł-a-ś*	-
3. 'He/she/it was'	*by-ł*	*by-ł-a*	*by-ł-o*
PL			
Person	MSC personal	Non-MSC personal	
1. 'We were'	*by-l-i-śmy*	*by-ł-y-śmy*	
2. 'You were'	*by-l-i-ście*	*by-ł-y-ście*	
3. 'They were'	*by-l-i*	*by-ł-y*	

In *Table 1* we can see that the past form of a verb is formed by adding *–l-* or *–ł-* to the root after deleting the infinitival ending *–ć*. This newly created form is called the *l*-participle. It is followed by gender markers *–e-*, *-a-* and *–o-* (for masculine, feminine and neuter gender respectively), *-i-* and *–y-* (for masculine personal and non-masculine personal gender), and by a person markers *–m, -ś-* and zero (for first, second and third person singular), *-śmy, -ście* and zero (for first, second and third person plural) (Sadowska 2012: 393).

The following paper consists of 6 sections in which I discuss two types of clitics in Polish: bound auxiliary clitics and free conditional morpheme *by*. In section 2 the diachronic development of Polish auxiliary clitics is presented as it is important to sketch the origin of modern clitic forms used in the contemporary language. Section 3 concentrates on the distribution of both kinds of clitics in Present-day Polish. In section 4, the Polish operator clitic *że* is introduced and discussed. Section 5 is devoted to the analysis of structural position of Polish auxiliary clitics. Section 6 is a short summary of the paper.

2. The diachronic development of Polish auxiliary clitics

In this section I will briefly sketch the origin of Polish past tense auxiliary clitics and conditional auxiliary clitic. We will have a look at verbal paradigms of past tense and conditional forms of the verb 'to be' which gave rise to the contemporary auxiliaries as clitics.

2.1. Past tense auxiliaries

What we call today 'mobile' or 'floating' past tense inflection arose as a consequence of a diachronic change from analytic to synthetic verbal structures, in particular the reanalysis of the auxiliary verb 'to be' as a person-number affix attached to the *l*-participle. Table 2 below presents the Old Polish orthotonic forms of the auxiliary verb 'to be' compared with Modern Polish forms of the copula verb 'to be' with the person-number agreement on each of them (Bąk 1984: 337):

Table 2: *The old and the present day paradigm of verb 'to be' in the present tense.*

Old Polish		**Present day Polish**	
SG	PL	SG	PL
1. *jeś-m*	1. *jes-my*	1. *jest-em*	1. *jest-eśmy*
2. *jeś-*	2. *jeś-cie*	2. *jest-eś*	2. *jest-eście*
3. *jest/ je/ jeść*	3. *są*	3. *jest*	3. *są*
1. DUAL *jeswa*		-	
2. DUAL *jesta*		-	
3. DUAL *jesta*		-	

The old paradigm was based on two different roots: *jeś-*, *jes-* and *są*, the second root occurred only in the third person plural. Interestingly, there was no ending in the second person singular. The dual forms disappeared in the 16th century, however, they can still be found in some dialects (Długosz-Kurbaczowa & Dubisz 2001: 305-308).

The Old Polish forms of the verb 'to be' in the present tense, presented in Table 2, and a verb in the past participle form with the thematic ending *-l-* or *-ł-* (so called *l*-participle) and gender inflection, formed the periphrastic past tense, which was the default form for speaking about past activities. As an example of the periphrastic construction, consider the data in Table 3 and Table 4 of the past tense for the verb *brać* 'to take' (Klemensiewicz et al. 1955: 367-373):

Table 3: *Periphrastic past tense forms for the verb brać – 'to take'.*

SG		
MSC	FEM	NT
1. *ja jeśm brał*	1. *ja jeśm brała*	-
2. *ty jeś brał*	2. *ty jeś brała*	-
3. *on jest brał*	3. *ona jest brała*	3. *ono jest brało*

Table 4: *Periphrastic past tense forms for the verb brać- 'to take'.*

PL	
MSC personal	Non-MSC personal
1. *my jesmy brali*	1. *my jesmy brały*
2. *wy jeście brali*	2. *wy jeście brały*
3. *oni są brali*	3. *one są brały*

The auxiliary could as well follow the past participle, as exemplified in Table 5 and 6 below:

Table 5: *Periphrastic past tense forms for the verb brać- 'to take'.*

SG		
MSC	FEM	NT
1. *ja brał jeśm*	1. *ja brała jeśm*	-
2. *ty brał jeś*	2. *ty brała jeś*	-
3. *on brał jest*[2]	3. *ona brała jest*	3. *ono brało jest*

[2] An anonymous reviewer wonders whether the pattern 'Subject-participle-auxiliary' was really attested in Old Polish texts because, as this reviewer points out, in the counterparts of such structures in South Slavic languages, the subject cannot be present if the participle precedes the auxiliary. The following are sentences from Old Polish biblical texts:

Table 6: *Periphrastic past tense forms for the verb brać- 'to take'.*

PL	
MSC personal	Non-MSC personal
1. *my brali jesmy*	1. *my brały jesmy*
2. *wy brali jeście*	2. *wy brały jeście*
3. *oni brali są*	3. *one brały są*

In 14th and 15th century both orderings of periphrastic past tense were alive (Klemensiewicz et al. 1955: 371):

(1) a. *wyszedł jeśm* *jeśm wywiódł*
I(masc) went out I(masc) led
b. *dała jeśm* *jeśm dała*
I(fem) gave I(fem) gave

Bąk (1984: 337) notes that at the beginning of the 14th century the auxiliary verbs *jest* in 3rd singular and *są* in 3rd plural start to disappear, and the finite forms *jeśm, jeś, jeśmy, jeście* begin to reduce their length (see data (2) to (6) below) and they ended up as what we call today 'movable' endings. Długosz-Kurbaczowa & Dubisz (2001: 307) also notice that the periphrastic past tense verbal forms started to evolve into shorter ones. They list the following processes which took part in the gradual change from the periphrastic past tense construction into the synthetic verbal forms which we have now. Consider also the following examples (2) to (6) which come from Długosz-Kurbaczowa & Dubisz (2001: 307).

1. In 3rd person singular and plural the auxiliaries *jest* and *są* started to disappear, what was left was only the *l*-participle.

(2) a. *uciekał jest* changed into: *uciekał*
'he ran away'
b. *uciekała jest* changed into: *uciekała*
'she ran away'
c. *uciekało jest* changed into: *uciekało*
'it ran away'

2. In 1st person singular the auxiliary was shortened to the form of *-eśm/ -śm* and it was attached to either the *l*-participle or some other word in a sentence, e.g.:

(3) a. *padł jeśm* changed into: *padłeśm*
'I (masc) fell down'

---

a) *I* ***uczynił jest*** *dzierżadła modrej barwy na kraju opony jednej na obu boku a na kraju opony drugiej takież* (BZ Ex36,11).

b) *…****wszedł jest*** *za mężem israelskim do domu niepoczesnego i* ***przekłoł*** *obu społu,męża *też i żonę, przez jich trzosła* (BZ Num 25, 8)

c) *…jichże to wszytko czysło* ***beło jest*** *dwadzieścia a dwa tysiącow a dwieście* (BZ Num 26,14).

In the examples (a) to (c) above we can see that the *l*-participle precedes the auxiliary in the sentence. In (a) and (b) the subject pronouns are omitted but in (c) the subject is overtly pronounced. The sentences in (a), (b) and (c) come from Słoboda (2012: 39, 50, 174).

b. *gdy jeśm szedł* changed into: *gdyśm szedł*
'when I (masc) walked'

3. In 2nd person singular the auxiliary was shortened to *-eś/-ś,* e.g.:

(4) a. *przysiągł jeś* changed into: *przysiągłeś*
'you (masc) promised'
b. *jakoż jeś osiadł* changed into: *jakożeś osiadł*
'as you (masc) settled'

4. In 1st person plural the auxiliary was shortened to *-smy,* which later became similar in pronunciation to the endings in 1st and 2nd person singular. Therefore, there was *-śmy* because of *-eśm/-śm* and *-eś/-ś*). Consider (10):

(5) *przyszli jesmy* changed into: 1. *przyszlismy – 2. przyszliśmy*
'we came'

5. In 2nd person plural the auxiliary was shortened to *-ście*, e.g.:

(6) *weseliły jeście* changed into: *weseliłyście*
'you (fem) enjoyed'

(Długosz-Kurbaczowa & Dubisz 2001: 307)

Długosz-Kurbaczowa & Dubisz (2001: 307) claim that the origin of the synthetic past tense verb form was due to the tendency for economy and simplification of the articulation effort as well as the need to fill in the gap with new synthetic constructions after the loss of the old simple past tenses (aorist and imperfectum).[3]

2.2. The development of conditional auxiliary by

The present-day Polish conditional mood auxiliary comes from an aorist form of the verb *być* 'to be'. Consider the following paradigm in (7) (Długosz-Kurbaczowa & Dubisz 2001: 315-316):

(7)

Old Polish aorist forms:		Present-day Polish conditional *by*	
SG	PL	SG	PL
1. *bych*	1. *bychom*	1. *bym*	1. *byśmy*
2. *by*	2. *byście*	2. *byś*	2. *byście*
3. *by*	3. *bychą*	3. *by*	3. *by*

The conditional mood construction was formed with the *l*-participle and the auxiliary *być* in the aorist form. The following examples (8)–(12) show the distribution of the conditional morpheme. It could either attach to the *l*-participle as in (8) and (11), precede it as in (9) and (10) or follow it as in (12):

[3] One of the anonymous reviewers points out that aorist and imperfectum were lost in Serbo-Croatian and Czech, too, but the auxiliaries were not reduced in those languages. So the reasoning in Długosz-Kurbaczowa & Dubisz (2001: 307) is only speculative. Another possible explanation given by this reviewer is the shift of lexical stress from the initial to the penultimate syllable in Polish (e.g. Migdalski 2006: 40).

(8) *zabiłbych* 'I (masc) would kill'
(9) *bych spadł* 'I (masc) would fall'
(10) *bychom zmarli* 'we would die'
(11) *opuścilibyście* 'you (pl.masc) would leave'
(12) *słyszeli bychą* 'they (masc) would hear'

(Długosz-Kurbaczowa & Dubisz 2001: 315-316)

Since the beginning of the 15th century, the aorist forms of the verb *być* started to resemble the Present-day Polish past tense form of this verb (specified in the brackets).

(13)

SG	PL
1. *bych --> bym (<--byłem)*	1. *bychom --> byśmy (<--byliśmy)*
2. *by --> byś (<--byłeś)*	2. *byście*
3. *by*	3. *bychą --> by(<--byli)*

(Długosz-Kurbaczowa & Dubisz 2001: 315-316)

Notice that in Present-day Polish the forms of 3rd person singular and 2nd person plural remained the same as in 15th century (consider data in (7) and (13).

3. Distribution of Polish auxiliary clitics

Polish has two types of auxiliary clitics: bound (used in a simple past tense) and a free morpheme *by* used to form the conditional mood. The paradigm for the perfect auxiliary clitics aligns with person-number inflection for the Modern Polish verb 'to be'. As shown in Table 7 below, the singular and plural morphemes for the 3rd person perfect auxiliary are morphologically null, and the 3rd person conditional morpheme does not have any person-number inflection. Conditional morpheme *by*, when encliticized to the *l*-participle, takes the auxiliary clitics to its right and they together form the conditional construction.

Table 7: *Polish past tense auxiliary clitics and conditional morpheme by.*

Past tense auxiliary		Conditional morpheme *-by*	
SG	PL	SG	PL
1. *-m*	1. *-śmy*	1. *by-m*	1. *by-śmy*
2. *-ś*	2. *-ście*	2. *by-ś*	2. *by-ście*
3. -	3. -	3. *by*	3. *by*

The following sections will analyse the synchronic distribution of the two kinds of Polish auxiliaries within a clause. In section 3.1 the syntax of perfect auxiliaries will be presented. Section 3.2 will deal with conditional morpheme *by*. Section 3.3 will show the differences in morpho-phonological behaviour between the auxiliary clitics and the *by* morpheme.

3.1. Mobile inflectional endings

Inflectional endings for present and past tense in Polish can attach to a variety of elements within the clause. However, the word that the inflectional ending is attached to has to precede the verb or to be the verb itself.

(14) a. *Ale się wyspała**m!***
but reflexive to sleep well.part.1st.sg.fem
'I slept so well!'
b. *Ale**m** się wyspała!*

In (14a) the past tense inflection is visible on the *l*-participle, which is the most common situation, whereas in (14b) it is attached to the conjunction *ale*, here used in a function of interjection. Similarly, in (15b) below the inflection is attached to a *wh*-word at the beginning of a question.

(15) a. *Kiedy wróci**łeś?***
when come back.part.2.sg.masc
'When did you come back?'
b. *Kiedy**ś** wrócił?*

The inflectional ending can also be separated from the copula verb in the present tense. In such a case, the rest of the verb is usually omitted.[4] Consider (16a) and (16b):

(16) a. *Głodny jeste**ś?***
hungry.nom.sg.masc be.pres.2.sg
'Are you hungry?'
*Głodny**ś?***

b.

SG	PL
1. *Głodny**m**.*	1. *Głodni**śmy**.*
2. *Głodny**ś**.*	2. *Głodni**ście**.*
3. -	3. -

(Bańko 2002:50)

In the following examples (17) – (24) different clausal elements have been shown to attract the past tense inflection.

(17) a. *Ty widziałe**ś**.*
you.sg see.part.2.sg.masc
'You saw.'
b. Ty**ś** widział. - subject pronoun

(18) *Ty jego**ś** widział.* - object pronoun
you.sg him see.part.2.sg.masc
'You saw him'

(19) a. *Ewy książkę czytałe**ś**.*
Ewa.gen.sg book.acc.fem.sg read.part.3.sg.masc
'You have read Eva's book.'
b. *Ewy-**ś** książkę czytał.* - inside the possessive DP

(20) a. *Janek powiedział, że pojechałe**ś** do Warszawy.*
Janek said that go.part.2.sg.masc to Warsaw. - complementizer *że*
b. *Janek powiedział, że**ś** pojechał do Warszawy.*

[4] This is true only for present tense copula verb 'to be' and only for first and second person singular and plural.

(21) *Tego chłopca**ś** lubiła.* - direct object (DP)
that boy like.part.2.sg.fem
'You liked that boy.'

(22) *Ty mamie**ś** kwiaty dał.* -indirect object (DP)
you.sg mother.dat.sg.fem flowers.nom.pl give.part.2.sg.masc
'You gave flowers to mother'

(23) *Daleko-**m** poszła.* -adverb (AP)
far walk.part.1.sg.fem
'I went far away.'

(24) *Świdomy**ś** był tych problemów.* -adjective (AP)
aware be.part.2.sg.masc these problems.gen.pl
'You were aware of these problems.'

Interestingly, some conjunctions and particles must take inflectional endings obligatorily.[5] Consider data in (25a-d):

(25) a. *Mama prosiła, aby**śmy** posprzątali.*
mother.nom.sg ask.part.3.sg.fem to tidy up.part.3.pl.masc
'Mother asked us to tidy up.'
b. **Mama prosiła, aby posprzątali**śmy**.*
c. *Czyżby**ś** zapomniał?*
question particle forget.part.2.sg.masc
'Is that so you forgot?'
d. **Czyżby zapomniał**eś?*

Bańko (2002: 51) notices that the attachment of inflectional endings to different elements in a clause is not completely free and is subject to some phonological (see example in (30)) or syntactic (see (29)) constraints. Sentences where the past tense inflection is separated from the verb are now most common in colloquial contexts. Some speakers may even have doubts whether that kind of constructions is grammatical. They may claim that they belong to some subdialect variety and not to a standard language. Bańko (2002) considers this a good judgement, since as he notices, elements which are remnants of archaic constructions are better preserved in a colloquial than in a standard language.

One of the most characteristic phenomena of Polish colloquial speech is adding the inflectional endings to the particle *że-*, as presented in (26a).

(26) a. *Dziewczynę **że**ś zranił.*
girl.acc.sg.fem that hurt.part.2.sg.masc
'You hurt a girl.'
b. **Dziewczynę że zraniłe**ś**.*
c. *Dziewczyn**ę**ś zranił. / Dziewczynę zraniłe**ś**.*

[5] The full list of these conjunctions is the following: *aby* 'for', 'in order to', *ażeby* 'in order to', *bodaj* 'perhaps', *bodajby* 'maybe', *bodajże* 'perhaps', *by* 'in order to', *byle* 'as long as', *byleby* 'so that', *chociażby* 'even if', *choćby* 'even if', *czyżby* 'is that so?', *gdyby* 'if', *iżby* 'that', *jakby* 'as if', *jakoby* 'as if', *jakżeby* 'so that', *jeśliby* 'if', *jeżeliby* 'if', *niechby* 'let', *niźliby* 'than', *niżby* 'than', *niżeliby* 'than', *oby* 'may, hope so', *toby, żeby* 'in order to' (Bańko 2002:50).

In (26c) we can see that the inflection can be attached to an object of a sentence or to a *l*-participle. If the particle *że-* is deprived of inflection, the sentence ends up ungrammatical, as shown in (26b). This is not true for a complementizer *że* ´that´. It can but does not have to carry the inflection. Consider (27a) and (27b):

(27) a. *Powiedziano mi, że**ś** przyjechał.*
b. *Powiedziano mi, że przyjechałe**ś**.*
'I was told that you came'

As was mentioned earlier, the distribution of floating inflections is not completely free. Although it may seem that they can be used everywhere in a clause, there are a few constraints. First of all, the word which carries the inflectional ending has to be the verb (participle) *or* to precede the verb in a clause. However, it does not have to precede it directly, as in (28a):

(28) a. *Daleko-**m** (od domu w ten piękny wiosenny dzień) poszła.*
'I went far away (from home on this beautiful spring day.)'
b. **Poszła daleko-**m**.*
'I went far away.'

In (28b) the sentence in not grammatical because the adverb which carries the inflectional ending follows the verb.

The inflectional ending cannot be attached to a transitive preposition, as in (29a) but it can be attached to a demonstrative pronoun, as in (29b):

(29) a. **Do-**ś** kina poszedłeś*
to cinema.gen.sg go.part.2.sg.masc
'You went to the cinema.'
b. T*ego-**ś** zadania nie wykonał.*
this task not do.part.2.sg.masc
'You did not do this task.'

In (30) below, a phonetic constraint disables the auxiliary clitic to attach to the previous word as there are two nasal sounds in immediate neighborhood which hinder the pronunciation of a string:

(30) **Ramię**m** zwichnął.*
arm.acc.sg.neut break.part.1.sg.masc
'I broke my arm.'

3.2. Conditional morpheme by + person-number agreement

When *-by* is affixed to the *l*-participle, the clause has the conditional (irrealis) meaning, similar to the meaning of English *would*. This morpheme can attach to the *l*-participle or stand on its own, as shown in (31a-b). However, when alone, it can only appear to the left of the *l*-participle in the clause. It carries the inflectional ending for person-number distinction.

(31) a. *Nigdy nie zrobiła**byś** tego.*
never neg.partic. do.part.2.sg.fem this
'You would never do this.'
b. *Nigdy **byś** tego nie zrobiła.*

Both in (31a) and (31b) the inflectional ending moves together with the *by* morpheme. The choice between these two variants depends mostly on prosodic factors. The *by* morpheme is an enclitic, so it attaches to the preceding word and it can neither occur at the beginning of a sentence, nor at the end. These restrictions do not apply to the conjunction *by* which can appear at the beginning of a sentence. Consider (32):

(32) ***By** tyle wiedzieć, trzeba dużo czytać*
that much know.inf. one should a lot read.inf.
'To know that much, one has to read a lot.'

The *by* morpheme, when it stands on its own, can appear in basically any place in a clause, except for the very first and the very last position. And the place right after the adverb seems to sound best for the sentence like in (33). Consider sentence (33) below:

(33) *(*By) Nasza Marysia (by) chętnie **by** dzisiaj poszła(by) do (*by) kina (*by).*
our Mary eagerly today go.part.3.sg.fem to the cinema
'Our Mary is keen on going to the cinema today.'

Also the distribution of the *l*-participle with the *by* morpheme affixed to it is considerably free. In this case the *l*-participle plus auxiliary may appear even at the very beginning and at the very end of a clause.

(34) *(Poszłaby)Nasza Marysia (poszłaby) chętnie dzisiaj poszłaby do kina (poszłaby).*
our Mary eagerly today go.part.3.sg.fem to the cinema
'Our Mary is keen on going to the cinema today.'

The distribution of the *l*-participle in a sentence without the morpheme *by* attached to it, is the same as with the morpheme by.

3.3. Auxiliary clitics and the morpheme by

Whereas the auxiliary clitics can attach to basically any element of any word category in a clause, the morpheme *by* does not exhibit that property. It can either stand on its own or be attached to the *l*-participle, and no other element.

(35) *Ty jutro(*byś) poszedł(byś) do kina(*byś).*
you tomorrow go.part.2.sg.masc to the cinema
'You would go to the cinema tomorrow.'

From (35) above we can see that the morpheme *by* cannot be attached to other element different than the *l*-participle. This is different from the properties of the auxiliary clitics, which when affixed to the word, constitute one phonological element with it.

The auxiliary clitics and the conditional morpheme *by*, when it stands on its own in the clause, behave similarly in the sense that they both can appear almost in every position in a clause. However, they are different with respect to what they can attach to – auxiliary clitics can attach to a wide range of elements, and the *by* morpheme can attach only to the *l*-participle.

4. Force-marking second position clitics

Second position clitics do not form a natural class in morpho-syntactic terms, as they include pronouns, adverbs, modals and sentential particles – the only property that unifies them is their prosodic deficiency and dependency. Operator clitics (Migdalski 2016), which are a special subgroup of Wackernagel clitics, are attested in all Slavic languages, irrespective of whether they have other second position clitics, verb-adjacent, weak pronouns or no other clitics at all. They mark Force-related contexts in languages like Bulgarian, Macedonian and Polish, which do not have other Wackernagel clitics.

The following section will deal with a distribution of the operator clitic *że* (subsection 4.1), its function and characteristics. Subsection 4.2 will present the auxiliary clitic *by* and its distribution in the subjunctive and conditional mood.

4.1. Operator clitic że

The operator clitic *że* is used to mark focus on the preceding word. Migdlaski (2009: 150) notices that the operator clitics are instances of the Wackernagel type of cliticization. He also suggests that they form a natural class by specifying the Illocutionary Force of a clause. Other examples of operator clitics are: Czech *prý*, which is used to report non-witnessed events; interrogative complementizer *li* (present in many contemporary Slavic languages such as Bulgarian, Macedonian, Russian, Serbo-Croatian); and operator clitic *že* in Czech and Russian.

In (36) below we can see the clitic *że* attached to the imperative verbal form. In this particular case it is used to mark stress on the element that precedes it.

(36) *Chodź**że** tutaj!*
come+Foc here
'Come here!'

(Migdalski 2016: 156)

In some cases, the clitic *że* may be used without the semantics of Focus for PF reasons. This situation happens when the clitic *że* appears with the auxiliary clitic on its right edge (Bański 2000, in Migdalski 2016: 156). The *że* insertion makes the sentence in (37b) grammatical because it enables the pronunciation of two nasal sounds which previously were in immediate neighbourhood as in (37a).

(37) a. **Ramię**m** zwichnął.*
arm.acc.sg.neut break.part.1.sg.masc
'I broke my arm.'
b. *Ramię-<u>że-</u>**m** zwichnął.*

Bański (2000) observes that *że* insertion can occur only when the auxiliary clitic needs to be prosodically supported, and it does not happen otherwise. So in a sentence such as (38c) it is possible to have double *że*. In this case, one of them is a complementizer and the other is a particle inserted for PF purposes (similarly to (37b) above). If the auxiliary clitic attaches to the *l*-participle and does not need any prosodic support, only one *że* may occur. Consider (38d) below:

(38) a. *Powiedziano mi, że**ś** przyjechał.*
b. *Powiedziano mi, że przyjechał**eś**.*
'I was told that you came.'
c. *Powiedziano mi, że że**ś** przyjechał.*
d. * *Powiedziano mi, że że przyjechał**eś**.*

In (38a) the auxiliary clitic attaches to the complementizer *że* in a subordinate clause. (38b) shows that the auxiliary clitic is attached to the *l*-participle which goes after the complementizer. In (38c) we can see a complementizer *że* and the operator clitic *że* to which the auxiliary is attached. The sentence in (38d) is ungrammatical because when the auxiliary attaches to the *l*-participle, only one że can appear, and this is the complementizer *że*.

4.2. Auxiliary by

The morpheme *by* when attached to the *l*-participle has the conditional (irrealis) meaning, similar to the meaning of English *would*. The affixation is not obligatory as *by* may precede the *l*-participle. The meaning and interpretation in such a case stay the same.

In a subordinate clause with the conditional or subjunctive mood the auxiliary is affixed to the complementizer and must occur in the second position in the embedded clause. This is similar to the structures with the operator clitic *że* in Polish, which also appears second. We have seen so far that Polish auxiliary clitics do not show second position cliticization patterns, the only exceptions are the operator clitic and auxiliary *by* in subordinate clauses (Migdalski 2016: 168). Consider the sentences in (39) below:

(39) a. *Chcę, że**byś** pożyczył mu książkę.*
want.1.sg.pres that lend.part.2.sg.masc him book
'I want you to lend him a book.'
b. *Chcę że pożyczył**byś** mu książkę.

(Migdalski 2016: 168)

Migdalski (2016) presents the data which shows that auxiliary clitics in Polish are second position clitics only in Force-marking contexts, these are contexts with subjunctive and conditional meaning used in subordinate clauses, like the one exemplified in (39a). Otherwise they are affixed to the *l*-participle.

5. The structural position of Polish auxiliary clitics

For the structural analysis I follow Veselovská and Emonds' (2015) model for IP projection where both the past tense and conditional auxiliaries occupy the I head with [-Modal] = Realis or [+Modal] = Irrealis features. The participle occupies the V head position.

Similarly to Veselovská and Emonds' (2015: 281-286) analysis of Czech, I assume that Polish past tense and conditional auxiliaries fit the conditions for being located in the head I. Like in Czech and English, Polish sentence negation particle *nie* occurs between I and VP. Consider the data in (40):

(40) a. Ja nie czekałam na autobus tamtego dnia.
I not wait.past part.sg.fem for a bus that day
'I was not waiting for a bus that day.'
b) Ja bym nie czekała na autobus tamtego dnia.
I aux+1.sg not wait.past part.sg.fem for a bus that day
'I would not wait for a bus that day.'
c) *Ja nie bym czekała na autobus tamtego dnia.
d) Ja czekałam nie na autobus tamtego dnia, ale na tramwaj.
'I was waiting not for a bus but for a tram.'

In (40c) we can see that the negative particle cannot occur before the conditional auxiliary *by*, and example (40d) shows that the sentence is acceptable but only as partial negation, and not as a sentence negation.

As it was presented earlier, the *by* morpheme, when it stands on its own, can appear in basically any place in a clause, except for the very first and the very last position. Consider:

(41) *Nasza Marysia (by) chętnie* ***by*** *dzisiaj poszła(by) do kina (*by).*
our Mary eagerly today go.part.3.sg.fem to the cinema
'Our Mary is keen on going to the cinema today.'

Also the distribution of the *l*-participle with the *by* morpheme affixed to it is considerably free. In this case the *l*-participle plus auxiliary may appear even at the very beginning and at the very end of a clause. See (42) below:

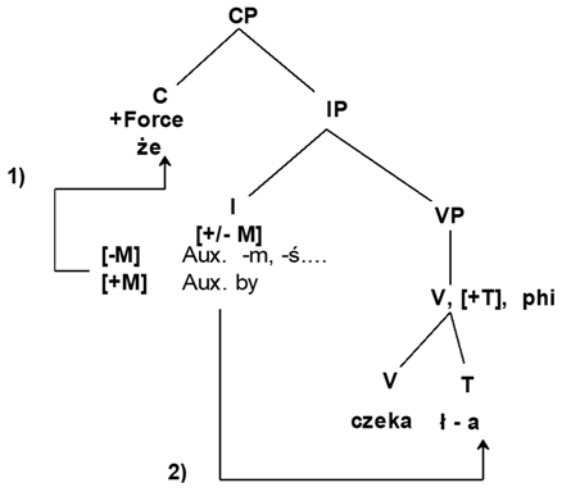
(42) *(Poszłaby)Nasza Marysia (poszłaby) chętnie dzisiaj poszłaby do kina (poszłaby).*
our Mary eagerly today go.part.3.sg.fem to the cinema
'Our Mary is keen on going to the cinema today.'

However, it is not possible for the conditional auxiliary *by* and the past tense auxiliary to occur in the initial position of a clause. See (43a) and (43b) below:

(43) a. **By /poszła nasza Marysia chętnie dzisiaj do kina (poszłaby).*
go.part.3.sg.fem our Mary eagerly today to the cinema
'Our Mary is keen on going to the cinema today.'
b. **śposzła chętnie dzisiaj do kina*
aux.2.sg, go.part.fem eagerly today to the cinema
'You were keen on going to the cinema today.'

Assuming that the position for both of the auxiliaries is in the head I, the past tense auxiliaries and the conditional morpheme *by* (when suffixed to a participle) move down in the structure and attach to the *l*-participle under V. This is marked by an arrow number (2) in the syntactic tree (44). The arrow number (1) shows movement whenever the Illocutionary Force related contexts appear (e.g. the conditional and subjunctive mood).

(44)



Below there are the two types of movement exemplified in sample sentences. Consider (45):

(45) M1. *Jan powiedział, żebyś czekała na niego w salonie.*
John said that.aux-aux.2.sg wait.part.sg.fem in the living room
'John said that you should wait for him in the living room.'
M2. Ty *czekałabyś na niego nawet sto lat.*
you wait.part.fem.aux-aux.2.sg for him even a hundred years
'You would wait for him even for a hundred years.'

In (45) the M1 sentence refers to the movement marked by arrow number 1 in (44). This is the movement of auxiliaries (in this case both of them: the auxiliary clitics and the auxiliary *by*) to the C head position motivated by the +Force feature where they attach to *że*. The other kind of movement marked by arrow number 2 in (44) is exemplified in the sentence M2. This is the movement of auxiliaries from I to V where they attach to the *l*-participle.

6. Summary

Polish auxiliary clitics do not show the Wackernagel second position behaviour in general. The only exception is the auxiliary clitic *by* in Force-marking contexts, so in the conditional and subjunctive mood, when they are encliticized to the particle *że* or some temporal adverbs, like *gdy* 'when' or clause initial conjunctions and complementizers (*jakby* 'as if', *oby* 'I wish that…'). The clausal negation and word order within a sentence allow to analyse the structural position of Polish auxiliaries as the head I of the IP projection.

References

Bańko, Mirosław. 2002. *Wykłady z polskiej fleksji*. Wydawnictwo Naukowe PWN, Warszawa.

Bański, Piotr. 2000. *Morphological and Prosodic Analysis of Auxiliary Clitics in Polish and English.* Doctoral dissertation, Warsaw University.

Bąk, Piotr. 1984. *Gramatyka Języka Polskiego*. Wiedza Powszechna, Warszawa.

Borsley, Robert and Maria-Luisa Rivero. 1994. Clitic Auxiliaries and Incorporation in Polish. *Natural Language and Linguistic Theory* 12: 373-422.

Długosz-Kurbaczowa, Krystyna and Stanisław Dubisz. 2001. *Gramatyka Historyczna Języka Polskiego*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

Embick, David. 1995. Mobile Inflections in Polish. *The Proceedings of the North East Linguistic Society* 25. Vol.2. Jill N. Beckman (ed.), 127-142. Amherst, MA. GLSA, University of Massachusetts, Amherst.

Klemensiewicz, Zenon, Tadeusz Lehr-Spławiński, Stanisław Urbańczyk. 1955. *Gramatyka historyczna języka polskiego.* Państwowe Wydawnictwo Naukowe. Warszawa.

Migdalski, Krzysztof. 2006. *The Syntax of Compound Tenses in Slavic*. Doctoral dissertation, Tilburg Unoversity. Utrecht: LOT Publications.

Migdalski, Krzysztof. 2007. On the emergence of the second position effect in Slavic. *Preceedings of FDSL 7, University of Leipzig, December 2007,* eds. Gerhild Zybatow, Uwe Junghanns, Denisa Lenertova, and Petr Biskup, 183-196. Frankfurt am Main: Peter Lang.

Migdalski, Krzysztof. 2009. On two types of Wackernagel cliticization in Slavic. In F*ormal Approaches to Slavic Linguistics: The Yale Meeting, 2008,* eds. Jodi Reich, Maria Babyonyshew, and Daria Kavitsakaya, 147-162. Ann Arbor, MI: Michigan Slavic Publications.

Migdalski, Krzysztof. 2010. On the relation between V2 and second position cliticization. *Lingua* 120 (2): 329-353.

Migdalski Krzysztof. 2016. *Second Position Effects in the Syntax of Germanic and Slavic Languages*. Wydawnictwo Uniwersytetu Wrocławskiego. Wrocław

Pancheva, Roumyana. 2005. The Rise and Fall of Second Position Clitics. *Natural Language & Linguistic Theory* Vol. 23, No. 1 (Feb., 2005), pp. 103-167.

Sadowska, Iwona. 2012. *Polish. A Comprehensive Grammar*. Routlege. New York.

Słoboda, Agnieszka. 2012. *Liczebnik w grupie nominalnej średniowiecznej polszczyzny, semantyka i składnia*. Wydawnictwo Rys, Poznań.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung, *Indogermanische Forschungen* 1, pp. 333–436.

Van Riemsdijk, Henk. 1999. Clitics: A state-of-the-art report. In Henk van Riemsdijk (ed.) *Clitics in the Languages of Europe*, Mouton de Gruyter, Berlin.

Veselovská, Ludmila and Joseph Emonds. 2015. The Cross-Linguistic Homes of Mood and Tense. In Henk van Riemsdijk and Mark van Oostendrop (eds.) *Representing Structure in Phonology and Syntax*. Berlin: Mouton de Gruyter, pp.277-314.

Spencer, Andrew. 2000. *Clitics. An Introduction.* Cambridge University Press, Cambridge.

Zwicky, Arnold. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.

LingBaW

Linguistics Beyond And Within 4 (2018), 79-91

Grammatical underpinnings of lexicalization patterns in Croatian, English and French: The case of [N PP] constructions

Daniela Katunar and Ida Raffaelli
University of Zagreb, Croatia

Abstract

This paper deals with the noun-preposition [N PP] construction in Croatian and compares the construction to its counterparts in English and French. Noun – preposition relations are analyzed as grammatical relations which participate in the formation of the lexicon, i.e. as grammatical devices which are productively used as *lexicalization patterns.* Based on the corpus analysis, [N PP] constructions in Croatian are identified and contrasted to English and French data. Lexical status of multi-word units in Croatian is discussed, as well as the level of idiomaticity of these constructions as compared to English and French. Whereas French and Croatian employ a similar lexicalization pattern, English uses compounding. The lexicon – grammar continuum is thus observed from the perspective of syntactic structures participating in word-formation.

Keywords: noun-preposition constructions, lexicalization patterns, Croatian, French, English

1. Introduction

Two processes typically associated with word-formation are derivation and compounding. This is especially common for analyses of morphologically rich languages such as Croatian, where various instances of multi-word units (e.g. *bijela kava* 'caffe latte; lit. white coffee', *kava s mlijekom* 'macchiato; lit. coffee with milk') are considered to be highly idiomatic, omitted from overviews in grammars (e.g. Silić & Pranjković, 2007) and generally viewed as belonging to the outskirts of the regular. Tafra (2005: 117- 119) points out that *lexicalization*, a process which in the broad sense refers to the creation of new words and in the more narrow sense to reanalyzing groups of words as words themselves (multi-word units), is as much a syntactic as it is a morphological (i.e. word-formation) challenge. In other languages, such as French or English, the relevance of multi-word units in structuring the lexicon has been studied to a greater extent. Booij (2009a: 222) discusses such instances in Germanic, Greek and Romance languages under the term *phrasal names,* which he defines as types of phrases which share the naming function

with complex words (2009a: 219). Similarly to Booij (2009a), Lyons (1995: 51) uses the term *phrasal expressions* when discussing lexically composite expressions and points out that such expressions are constructed using grammatical rules of a language. These rules include both morphological and syntactic devices used in the formation of the lexicon. Different grammatical devices that participate in the formation of the lexicon emphasize the long-term problem of uniformly defining the notion of the 'word' in linguistic theory, which is present from structuralist discussions onwards.[1] Also, this problem is what drives contemporary theories of grammar, such as Construction Grammar, to argue for a continuum between lexicon and grammar, in large part based on the fact that grammatical rules or schemas play a significant role in building the lexicon, traditionally regarded as a repository of idiosyncratic forms. In this study we take the basic tenets of Construction Grammar (Goldberg 1995, 2006) as the theoretical and methodological background against which the [N PP] construction is analyzed. More specifically, we consider the [N PP] construction as an instance of a more general grammatical relation (noun post-modification), but with a specific set of syntactic and semantic features which are pertinent to the process of lexicalization.

Thus the goal of this study is to investigate in greater detail the [N PP] construction[2] in Croatian, e.g. Cro. *četkica za zube* 'toothbrush', *odbojka na pijesku:Loc* 'beach volleyball', *alergija na mačke* 'cat allergy', and contrast it to two typologically and genetically different languages – English and French. As we will show in our data, English and French differ with respect to their lexicalization strategies, i.e. productive morphosyntactic patterns used in comparison to Croatian. Most notably, English uses compounding as a productive grammatical device in word-formation, whereas French [N PP] constructions are considered to be a very productive lexicalization pattern. Instead of the term *phrasal expression*, we use the term [N PP] construction in order to emphasize the grammatical underpinnings of its lexical formation, in line with the aforementioned emphasis on the lexicon-grammar continuum.[3] Moreover, the term points to one of its main properties of this type of construction. It is a particular lexicalization pattern with a specific naming function consistent with the criteria offered by Lyons (1995) and Booij (2009a).

There is little extant literature on the construction in Croatian, although generally authors view noun-preposition expressions as a grammatical phenomenon related to either noun modification (Silić & Pranjković, 2007, Belaj & Tanacković Faletar, 2014), complementation (Petrović 2016) or collocations (Blagus Bartolec 2014). The notion of construction used in this paper is broader than that of collocations because it places emphasis on the productivity of the syntactic pattern, as well as specific semantic and symbolic links that hold between nouns in the [N PP] construction. Furthermore, the emphasis is placed on the semantic (and/or functional) contribution of the preposition, a point that is omitted in extant overviews.

Raffaelli (2015) emphasizes the tendency of [N PP] patterns in Croatian towards lexicalization, especially drawing parallels to French examples, e.g. *pomme de terre* 'potato'. In

---

[1] For example see Martinet's influential paper *Le Mot* (1966).

[2] Note that this is different from the [N P N] construction as defined by Jackendoff (2008), e.g. *day by day, dollar for dollar, book upon book*.

[3] This is also in line with Booij's (2009a: 237) notion of a constructional schema in his analysis of A+N phrases with a naming function.

this respect the [N PP] construction can be regarded as a type of *lexicalization pattern* i.e. a relation between meaning and surface expression, with the potential of being a typologically relevant pattern productive in the language (Talmy, 1985). The term lexicalization pattern is used mostly in connection to Talmy's work on the typology of lexicalization patterns of motion and location verbs (1985). The concept has been further expanded in Raffaelli and Kerovec (2017) to include word-formation patterns and other syntactic processes that are used in lexicalization of particular domains. Therefore, in the broad sense we consider as lexicalization patterns any arrangement of formal morphosyntactic elements with a specific meaning or function which can be observed contrastively.

The typological relevance of investigating this construction in Croatian is the motivation behind the contrastive analysis with English and French. We believe that such a tripartite contrastive analysis can have broader implications in examining grammatical processes in three large Indo-European language families – Slavic, Germanic and Romance – that represent a kind of regular and frequent lexicalization pattern used in naming certain semantic domains.

The main questions we will address in the paper are:

a) is the [N PP] construction productive in Croatian and what is its role in a morphologically rich language such as Croatian,
b) what lexicalization strategies are used in English and French as compared to Croatian as the source language,
c) what can a contrastive analysis reveal about the potential lexical status of Croatian [N PP] constructions,

With respect to lexicalization patterns, we must emphasize that we will use Croatian as a source language in our contrastive analysis. This means that we will expect different lexicalization patterns to emerge from our comparison to French and especially English, because it uses compounding more extensively than [N PP] constructions.

2. Noun – preposition relations in Croatian, English and French

When it comes to investigating to role of prepositions in syntax and semantics, noun–preposition relations (of the [N PP] type) can be viewed as secondary to verb–preposition relations, which are more often the focus of syntactic-semantic analyses in literature (e.g. Fillmore, 1968, 1977; Jolly, 1991; Rauh, 1993; Paillard, 2002; Katunar, 2015, Katunar & Raffaelli, 2018). This is based on the view that in many languages without morphological cases, such as English and French, prepositions fill the role of cases in the argument structure of the sentence, with the verb at its centre. In syntactic-semantic analyses of prepositions in Croatian, prepositions are also commonly discussed in relation to case semantics, and are viewed as governing particular cases within the preposition phrase [PP]. For example, the expressions *na stolu.Loc* 'on the table', *na stol.Acc* 'on the table', *sa stola.Gen* 'off of/from the table' are defined as preposition–case constructions with the meaning of Location, Goal and Source, respectively. The prepositions *na* 'on' and *s(a)* 'off (of)' add the spatial meaning of 'two-dimensional surface', in addition to the semantic roles coded by the entire preposition-case construction. Such analyses fall in line with the localist theories of cases and are also expanded by cognitive

linguistic analyses of preposition semantics which view spatial meanings as the basis for extensions of non-spatial meanings as well, e.g. *na Božić* 'at Christmas'. [4] However, these analyses often also rely on the broader sentence context to assess the meanings of preposition-case constructions, such as those in (1) and (2), which is also centered around the verb and its argument structure, *biti* 'to be+location', *staviti* 'to put+goal'.

(1) *Knjiga je na stolu.*
'the book is on the table'

(2) *Stavio sam knjigu na stol.*
'I put the book on the table'

When it comes to noun-preposition relations, these are often analyzed as cases of noun modification or noun complementation, both processes being grammatical and not related to lexicalization. However, the distinction between complements and modifiers is not straightforward in the extant literature. In Croatian grammars (Silić & Pranjković, 2007: 265) preposition phrases that follow nouns are modification phrases if they perform an attributive function, e.g. *udžbenik na hrvatskom jeziku* 'a textbook in Croatian', *ples po kiši* 'a dance in the rain', *putovanje preko granice* 'a voyage across the border'. They are considered atypical forms of modification because they depend on the relation of government instead of the typical relation of agreement present in the modifying Adjective-Noun constructions, e.g. *mirno more* 'calm sea'. Furthermore, examples of post-modification are usually facultative and semantically transparent. For instance, Belaj and Tanacković Faletar (2014: 184-186) give the following expression in (3) as an example of post-nominal modification:

(3) *mala lopta ispod stola.Gen*
'small ball under the table'

This pattern is interchangeable with other post-nominal modification phrases, e.g. *mala lopta crvene boje.Gen* 'small ball of red color', *mala lopta iznad/preko puta stola* 'small ball over/across from the table', i.e. it is fully transparent semantically and allows for facultative variability common for a productive grammatical schema. This is very similar to the way preposition phrases are viewed in English grammars, with PP being the most frequent post-modifier element in complex noun phrases (Biber et al., 1999: 606).

On the other hand, prepositional complements to nouns in Croatian[5] are also discussed by Petrović (2016) as being facultative attributes with (mostly) weak government between the noun and preposition phrase, e.g. *pozivnica za Tanju* 'an invitation for Tanja', *zaštita od/protiv hladnoće* 'protection from/against the cold'. Another type of preposition complementation is given in Katunar (2015) and includes strong government which is often transferred from a specific verb-preposition relation, e.g. *vjera u čovječanstvo* 'belief in humanity' <*vjerovati u čovječanstvo* 'to believe in humanity'. Due to the relation of strong government these constructions are not facultative, but also do not conform to the criteria of naming function.

[4] See Pranjković (2002), Kerovec (2012), Belaj & Tanacković Faletar (2014), Katunar (2015) and references therein.

[5] As compared to German.

Therefore, it is clear that the notions of government, complementation and modification are not strictly separated when it comes to many instances of noun-preposition relations. More importantly, the question of lexicalization of some of these expressions is not raised in such discussions.

However, it should be pointed out that in French linguistics Martinet (1989) analyzed with scrutiny the noun-preposition relation as a pattern used in formation of new linguistic expressions. It is well known that Martinet was an arduous critic of the notion and the term *word* in linguistics because of its inconsistency and theoretical incoherence. He considered it as not specific enough, especially when defined as a string of phonemes between two blanks on the paper. His comparison of the English lexeme[6] *potato* and the French lexeme *pomme de terre* is well known. According to Martinet (1989), these linguistic units should both be considered autonomous syntagms. They both refer to the same referent, they share the same meaning, but in a structurally different manner. The lexeme *pomme de terre* is a linguistic unit representing a coherent syntagmatic structure and encoding a certain meaning. If one component (linguistic element) within the N PP structure would be replaced with another, e.g. the PP *de terre* 'lit. of the earth' with the adjective *rouge* 'red', the meaning of the syntagm *pomme de terre* would completely change. Instead of potatoes, we would have red apples (*pomme rouge*). Martinet argues for the term *autonomous syntagm* because it refers to different types of syntagmatic structures representing a coherent linguistic unit with respect to form and meaning. According to its typological features each language uses different linguistic elements and structures to form an autonomous syntagm. For example Croatian *automobil**om***, French ***en*** *voiture* and English ***by*** *car* represent equal autonomous syntagms, but with different types of functional morphemes used. Croatian in this respect significantly differs from French and English, although the three expressions share the same meaning. Croatian uses the flective morpheme *–om* (encoding instrumental case), whereas French and English use prepositions *en* and *by*. For linguists it is important to recognize differences and similarities in linguistic elements used in structuring autonomous syntagms. It should be mentioned that Martinet already points out that linguistic units of the N PP type like *brosse à dents, machine à laver* represent a kind of classification of entities.

Therefore, what we consider to be the defining feature of lexicalization in the case of [N PP] constructions is the naming function typical of phrasal names, e.g. *gel za tuširanje* 'shower gel', *daska za rezanje* 'chopping board', *hokej na travi* 'field hockey', *čaj od mente* 'mint tea'. The naming function should allow for an [N PP] expression to denote a class of entities, and not simply perform the function of a referential grounding of expression, as in (3) *mala lopta ispod stola* 'a small ball under the table'. Therefore, neither the complementation nor post-modification examples in the extant literature on Croatian are analyzed with regards to the naming function typical of phrasal names as part of the lexicon.

[6] Note that Martinet (1989) does not use the term lexeme in the same sense as it is commonly used with respect to Anglo-Saxon tradition (see Raffaelli 2015).

3. Identifying [N PP] constructions in Croatian

Lexicographic resources in Croatian offer scarce data on the [N PP] construction and quite often do not list any of its instances as either separate lexical entries or possible idioms or collocations. This is unlike many multi-word units of the form Adjective-Noun, e.g. *bijela kava* 'caffe latte', *crna rupa* 'black hole', which are often listed as collocations (or syntagms). For instance, the entry *četkica* (*četkica za zube* 'toothbrush') is listed in two main monolingual dictionaries of Croatian, the online Croatian Language Portal[7] and the *Dictionary of Croatian Language* (Šonje, 2000), only as a diminutive form of *četka* 'brush'. Therefore, in order to extract candidates for [N PP] constructions in Croatian we used the Croatian Web Corpus HrWaC, the largest general corpus of Croatian containing 1,9 billion tokens. Apart from some inconsistencies expected in large general dictionaries, the fact that the examples we extracted from corpora are not listed in those dictionaries may point to two important properties of the [N PP] construction. Firstly, it is behaving as an emergent lexicalization pattern and secondly, though the choice of the preposition phrase is fixed, it is omissible in some contexts (Raffaelli, 2015). For example, the following sentence in (4) is ambiguous without context.

(4) ? Imamo različite vrste lakova
'We have various types of polishes'

However, in the extralinguistic context of a beauty parlor or a hardware store the noun *lakovi* 'polishes' could easily be inferred by the speaker to mean either *lak za nokte* 'nail polish' or *lak za drvo* 'wood polish', respectively. On the other hand, some expressions also vary in corpora between two types of modification patterns – [A N] and [N PP], e.g. *dječji doplatak – doplatak za djecu* 'childcare allowance, childcare benefit'. *Dječji doplatak* has 1929 occurrences in HrWaC), and *doplatak za djecu* 600 occurrences. Though the frequencies surely point to the [A N] pattern as the dominant one, it is important to point out that in legislative documents (laws, administrative texts and official newspaper reports), i.e. those texts that define and control childcare benefits, the pattern is [N PP]. [A N] is mostly used in unofficial texts. In such cases it is important to note that corpus frequencies of particular [N PP] constructions may be skewed due to competing patterns. In many cases, however, it is not possible to use [A N] patterns at all, e.g. **noktni lak* 'nail polish', **zubna četkica* 'toothbrush'. Furthermore, in other cases [A N] carries a different meaning and connotation than [N PP]. For example, *azil za životinje* means 'animal shelter. lit. asylum for animals', whereas *životinjski azil* 'lit. animalistic asylum' would carry the meaning 'animalistic, aggressive, crude' type of shelter.

Four of the most frequent primary prepositions in Croatian were chosen for the analysis[8], *na + Locative/Accusative* 'on'*, za + Accusative/Instrumental* 'for'*, od + Genitive* 'from' and *s(a) + Genitive/Instrumental* 'off of / with'.[9] Although prepositions are very frequent and

[7] http://hjp.znanje.hr/

[8] Though other primary and secondary prepositions were queried as well, e.g. *ispod* 'below', *iznad* 'above' etc.

[9] Two other frequent morphologically simple prepositions were checked, *u + Accusative/Locative* 'in' and *o + Accusative/Locative* 'about'. These prepositions, however, did not yield any significant number of [N PP] constructions in our data.

polysemous units, they can be regarded as the constant in the [N PP] construction when it comes to corpus queries. We searched for all [N P N] sequences, and then manually extracted cases of [N PP] constructions into a database based on our criteria mentioned above – naming a separate class of entities or phenomena.

The structure of the [N PP] construction allows for some variation within the PP of two less frequent forms, [N P NN_{GEN}], e.g. *aparat za gašenje požara.Gen* 'fire extinguisher', and [N P [N i N]], e.g. *dom za starije i nemoćne* 'retirement home; lit. home for the elderly and feeble'. The latter instance is only found in the example 'retirement home'.

Table 1. below lists the number of different [N PP] constructions for each of the prepositions + cases we extracted from HrWaC. From a quantitative point of view, the frequencies would suggest it is a secondary lexicalization pattern when it comes to multi-word units, but still productive across various instances. However, the notion of the conventionalization of this pattern cannot rely solely on corpus frequency data, especially taking into account the types of entities being named, which are everyday basic objects, as we will show in the next section. Table 1. also shows the order of productivity of particular prepositions.

Table 1: *Number of total occurences of [N PP] construction with a particular preposition extracted from HrWaC.*

Preposition + case	**Number of [N PP]**
od + genitive	769
za + accusative	446
na + accusative	114
s(a) + genitive	113
na + locative	58
s (a) + instrumental	21
za + instrumental	4

4. Some properties of [N PP] constructions in Croatian – domains, schematicity and specificity

In order to get an insight into the types of entities being lexicalized by the [N PP] construction, i.e. its classificatory properties, we divided our data into domains. These domains are:

1) ARTEFACTS: *četkica za zube* 'toothbrush', *aparat za kavu* 'coffee machine', *vreća za smeće* 'garbage bag', *perilica za rublje* 'washing machine', *palica za golf* 'golf club', *palica za bejzbol* 'baseball bat', *papir za pečenje* 'parchment paper, baking paper', *polica za knjige* 'bookshelf', *čamac na vesla* 'rowboat', *ekran na dodir* 'touch screen', *dvorac od pijeska* 'sand castle', *škrinja s blagom* 'treasure chest', *ormarić s lijekovima* 'medicine cabinet'
2) FOODSTUFF: *salata od piletine* 'chicken salad', *čaj od mente* 'mint tea', *vrhnje za kuhanje* 'heavy cream, cooking cream', *sirup od bazge* 'elderflower syrup', *dagnje na buzaru* 'mussells stew'
3) INSTITUTIONS: *dom za djecu* 'children's home, orphanage', *agencija za zaštitu okoliša* 'environmental agency', *odjel za istraživanje* 'department of research', *klinika za traumatologiju* 'trauma clinic'
4) ECONOMY and LAW: *porez na dohodak* 'income tax', *jamac za kredit* 'loan guarantor, loan cosigner', *dozvola za rad* 'work permit', *doplatak za djecu* 'childcare allowance'
5) MEDICINE: *kapi za uši / oči / nos* 'ear /eye / nose drops', *sirup za kašalj* 'cough syrup'

6) SPORTS: *odbojka na pijesku* 'beach volleyball', *četverac / dvojac na pariće* 'quad / double scull', *ronjenje na dah* 'freediving'
7) SOCIO-CULTURAL RELATIONS: *veza na daljinu* 'long-distance relationship', *lov na vještice* 'witch hunt'
8) TITLES: *Albert od Monaka* 'Albert of Monaco'

The examples provided in these domains point to a notion of productivity not related to corpus frequency, but still relevant when it comes to discussing the saliency of [N PP] lexicalization pattern – the entities named are common objects, terms and activities which are found in everyday life and are culturally salient. Various artefacts – grooming implements, home appliances, foodstuff and medicines are some of the most basic items one can encounter. The domains we defined are of course broad domains of experience in general, and the [N PP] construction performs two specific functions when it comes to structuring parts of the lexicon: a) specialization of subordinate (hyponymy) members of a category and b) specification of polysemous or vague nouns.

a) The first function relies on the pronounced schematicity of this construction that allows for novel expressions to be made productively. It behaves as a periphrastic expression, being more semantically transparent than opaque. This is in part due to its grammatical form, but also it can be tied to what is known in Construction Grammar as an *encoding idiom*, an idiom whose meaning can be inferred from its parts but the speaker still would not know if it is the conventional way of saying something (Goldberg, 1995: 155).[10] Goldberg gives the example of *sofa bed* which is in Croatian translated as *kauč na razvlačenje.Acc* and thus fits the description of encoding idioms quite well.

The schematicity in part accounts for the high frequency of some of the preposition + case combinations. For example, a significant part of [N PP] examples with the prepositions *od + genitive* refer to various sorts of food and drinks, and allow for specification of certain types of foodstuff via the prepositional phrase, as in (6), where various types of herbal tea are named by the same subpattern *čaj od* __.

(5) čaj od __: borovnice / mente / bazge / jagode / kamilice / koprive…
'blueberry / mint / elderflower / strawberry / camomile / nettle… tea'

Such instances are usually subordinate members of a category (e.g. tea), and are present as hyponyms in the taxonomy. Therefore, [N PP] construction has a specialization effect of creating various co-hyponyms on the lower levels of taxonomy. This is pronounced in other examples as well, like many government agencies, kinds of taxes (income, property, estate tax and so forth). Interestingly, in higher levels of taxonomies, the [A N] pattern is usually the dominant one, as in Figure 1:

[10] As Goldberg notes, this notion was first put forth by Makkai (1972).

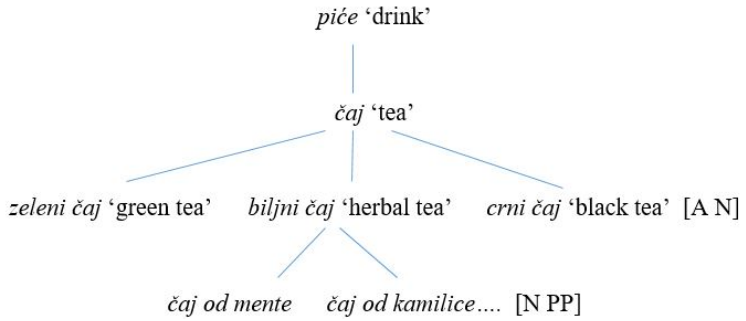


Figure 1: *Lower levels of a taxonomy where [N PP] construction specifies "types of".*

b) Secondly, there are many head nouns in our data which exhibit a type of vagueness. For example, *daska* 'board' can refer to a flat object, prototypically made from wood. However, the expressions *daska za jedrenje* 'windsurf board', *daska za rezanje* 'chopping board', *daska za peglanje* 'ironing board' all specify different objects with different functions and thus resolve the ambiguity. However, they cannot be treated as co-hyponyms because they belong to different categories – sports equipment, cooking tools, home equipment. Other nouns behave similarly, e.g. *aparat za kavu* 'coffee machine', *aparat za dijalizu* 'dialysis machine', *aparat za gašenje požara* 'fire extinguisher' – they all denote different types of machines with different functions.

The [N PP] construction can thus be analyzed as playing a part in two crucial areas of lexical structure – that of hierarchy relations and polysemy relations.

5. Contrastive analysis of French and English to Croatian

As we mentioned in previous sections, French and English use mutually different lexicalization strategies. In French there is a long tradition of regarding [N PP] constructions as an important part of the lexicon. In English, compounding is not reserved only for fixed expressions but is used productively and it is considered to form a subsystem of grammar as well (Jackendoff, 2009). In this way compounding is a way to produce lexically composite and novel expressions. The two languages examined together in comparison to Croatian reveal some interesting insights into the level of idiomatization of Croatian [N PP] constructions as well as their claim to lexical status.

The contrastive analysis was performed with Croatian data as the starting point. This means that [N PP] constructions collected from the Croatian corpus were translated into French and English and then particular patterns were examined.

5.1. English and Croatian

When it comes to English, the lexicalization strategy is predominantly compounding. The correlation between the lexicalization patterns is that the head noun N1 in Croatian corresponds to head noun N2 in English compounds, Cro. [N1 P N2] – Eng. [N2 N1][11], e.g.

[11] We use the term *head* in line with the traditional definition that the head determines the syntactic and semantic category of a compound (Booij, 2009b), and in the case of Croatian, the syntactic and semantic category of the [N PP] construction.

vreća:1 za smeće:2 – garbage:2 bag:1, lak:1 za nokte:2 – nail:2 polish:1. The headedness of the two patterns thus reveals a grammatical consistency as well, both in word order and the semantics of the expressions. Furthermore, as English counterparts to Croatian examples can be found in dictionaries of English, such as the *OED* and *Merriam Webster Online Dictionary*, it is clear that Croatian [N PP] constructions acquire a lexical status of a sort. Interestingly, in the *English-Croatian or Serbian Dictionary* (Filipović, 1987) lexemes such as *tooth-brush, nail polish* and *tooth-paste* are translated as *četkica za zube, lak za nokte* and *pasta za zube.*

Another relevant observation relates to the status of prepositions in Croatian. While prepositions are considered to be on the border between lexical and functional words, it seems that their role in [N PP] constructions is highly grammaticalized and lacks the prototypical spatial semantics of these primary prepositions. For instance, the preposition *na+Accusative* profiles mainly MANNER, e.g. *jaje na oko* 'egg sunny-side up', *čamac na vesla* 'rowboat', *ronjenje na dah* 'freediving'*, lutka na napuhavanje* 'love doll'. The preposition *za+Accusative* profiles mainly PURPOSE, e.g. *daska za peglanje* 'ironing board', *gumica za brisanje* 'eraser', *pjena za brijanje* 'shaving cream', UNDERGOER, e.g. *perilica za rublje* 'washing machine; lit. washer of laundry', *lak za nokte* 'nail polish' and BENEFICIARY *kućica za psa* 'doghouse', *azil za životinje* 'animal shelter'.[12] *Od+genitive* 'from' is used for SOURCE MATERIAL, *dvorac od pijeska* 'sand castle', while *s(a)+instrumental* 'with' is used for containment/content for inanimates (*škrinja s blagom* 'treasure chest'). *Na+locative* is the only element which still holds the spatial meaning in PPs, e.g. *kućica na drvetu* 'tree house, lit. house on the tree', but it is being reanalyzed as the integral part of the name of the entity. These roles are what the preposition phrase explicitly adds to an expression and are useful for naming properties which would not be explicitly named in cases of compounding or [A N] patterns. In English, on the contrary, such roles are usually inferred by the speakers from the compound form itself.

A few exceptions have to be noted as well. These have to do with two domains where instances of preposition phrases are found in the English data as well – institutions and titles, e.g. *odsjek za lingvistiku – department of linguistics*, *centar za autizam – center for autism, Albert od Monaka – Albert of Monaco.* However, these examples are scarce in the data and there are counterexamples in the same domains, e.g. *klinika za traumatologiju – trauma clinic, trauma center*. One correlation is with the descriptive *of*-construction and *za+accusative*. Another, between *for* and *za+accusative* may be skewed due to Croatian being the starting point because some of the institutions are culturally specific and the English translation is adjusted to the original name in Croatian, e.g. *agencija za mobilnost – agency for mobility.* Another systematic correlation is with prepositions *with* and *s(a)+instrumental.* However, this is found in the subdomain of people with medical conditions, e.g. *osobe s invaliditetom – people with disabilities*, and may have to do with the distancing the condition of the subject that the preposition *with-s(a)+instrumental* offers, for reasons of politeness and correctness.

English, therefore, with an abundance of compounds serving as equivalents to Croatian [N PP] across all domains provides a good cross-linguistic argument for treating Croatian [N PP] constructions as instances of complex words, i.e. of phrasal names.

---

[12] Some of these are quite regular. For instance, *za+Accusative* with PURPOSE usually takes the gerund forms, i.e. deverbal nouns with the suffix *–nje* (*jedrenje* 'windsurfing', *rezanje* 'cutting').

5.2. French and Croatian

When it comes to French, the N PP constructions are predominantly used in correspondence to Croatian N PP examples. However, there are some examples where French uses simple lexical units whit respect to Croatian N PP constructions such as *cafetière : aparat za kavu* 'coffee maker'; *lessive: prašak za rublje* 'detergent'*; poubelle : koš za smeće* 'trash bin'. This difference is mostly related to the artefact domain, with respect to which Croatian is very productive in using N PP constructions as a lexicalization pattern.

It should be pointed out that the two most frequently used prepositions in the French N PP constructions are *de* and *à*. They are highly grammaticalized and therefore idiosyncratic, appearing in constructions referring to all the domains (except medicine), as defined for the Croatian N PP constructions. These are: artefacts – *brosse à dents* 'toothbrush', institutions – *maison de retraite* 'retirement home' economy – *permis de travail* 'work permit', food – *agneau à la broche* 'lamb on a spit', sport – *planche à voile* 'sailboard', titles – Albert de Monaco. Regarding the domain of medicine, two prepositions that are systematically used in the French PP constructions are *contre* and *pour*, as in *remède contre la toux / cancer / SIDA* 'cough syrop / cure for cancer / AIDS', *gouttes pour le nez* 'nose drops'. In Croatian there is also a competing pattern with two antonymous prepositions: *za+accusative* 'for' and *protiv+genitive* 'against', used in the domain of medicine, e.g. *sirup protiv kašlja – sirup za kašalj* 'lit. syrup against/for the cough'. However, when the frequencies in HrWac were checked, *sirup za N* is twice as frequent than *sirup protiv N* (382 versus 183 occurrences), for example. This shows that the purpose-oriented meanings of *za+accusative* generalized in the [N PP] construction.

Firstly, when comparing Croatian and French usage of the N PP constructions as kind of lexicalization pattern, one should note that the two languages exhibit a high degree of consistency in using the N PP constructions to lexicalize the same domains. Secondly, the French preposition *à* corresponds to the usage of all the basic primary prepositions in the Croatian N PP constructions, such as *na:Acc – lov na vještice – chasse aux sorcières* 'witch hunt'; *za:Acc – perilica za rublje – machine à laver* 'washing machine'; *na:Loc – janjac na žaru – agneau à la broche* 'lamb on a spit'; *s(a):Instr – škrinjica s blagom – coffre au trésor* 'tresaure chest'. Therefore, *de* and *à* are more grammaticalized and idiosyncratic than Croatian PPs in the sense that the particular semantic import from the preposition is not as predictable as it is in Croatian. However, it should be pointed out that *od:Gen* is the only Croatian primary preposition which systematically corresponds to the French preposition *de*: *kora od banane- poile de banane* 'banana peel'. When the PP refers to a material of which a certain item is made of as in *haljina od svile* 'silk dress'*, nakit od srebra* 'silver jewellery', the French preposition *en* is systematically used as *in robe en soie, bijoux en argent*.

Furthermore, the comparative analysis of Croatian and French N PP constructions shows that in both languages the N PP constructions cover the same range of semantic domains, and thus should be considered as a frequent and highly entrenched lexicalization pattern. It means that the cross-linguistic analysis of the same grammatical structure – the N PP construction – could point to some frequent and regular grammatically underpinned strategies speakers use in lexicalization, i.e. naming certain items. From the lexical-semantic point of view, these

constructions share the same classification principle. They all represent hyponyms within taxonomic lexical structures, which also points to their highly corresponding regular features.

6. Conclusion

The results of the analysis of [N PP] constructions in Croatian as compared to English and French reveal a grammatically based noun-preposition relation being used in the service of lexicalization. The main properties of the [N PP] construction, specialization of hyponymy relations and specification of vagueness, make it an essential part of lower levels of the lexicon hierarchy, both in Croatian and in French. The contrastive analysis showed that English and French differ systematically in their lexicalization strategies, with English using compounding extensively in the place of [N PP]. Moreover, the analysis shows that Croatian and French frequently and regularly use the corresponding grammatically underpinned strategy to lexicalize the same semantic domains. The analysis points to the model of lexicalization patterns as typologically relevant capturing diverse grammatical structures being productive in the process of lexicalization. Such an examination of corresponding lexicalization patterns inter- and intra-linguistically lays the ground for future comparisons and establishes a theoretical framework for examining the relationship between lexicalization, grammaticalization and idiomatization.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). Cambridge, MA: MIT Press.

Blagus Bartolec, G. (2014). *Riječi i njihovi susjedi. Kolokacijske sveze u hrvatskom jeziku.* Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Booij, G. (2009a). Phrasal names: A constructionist analysis. *Word Structure*, *2*(2), 219-240.

Booij, G. (2009b). Compounding and construction morphology. In R. Lieber & P. Stekauer (Eds.), *The Oxford Handbook of Compounding* (201-216). Oxford: Oxford University Press.

Belaj, B., Tanacković Faletar, G. (2014). *Kognitivna gramatika hrvatskoga jezika, Knjiga prva, Imenska sintagma i sintaksa padeža*. Zagreb: Disput.

Fillmore, Ch. J. (1968). The case for case. In E. Bach, Th. Harms (Eds.), *Universals in Linguistic Theory* (1-88). New York: Holt, Rinehart, and Winston.

Fillmore, Ch. J. (1977b). The case for case reopened. *Syntax and semantics*(8), 59-82.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Jackendoff, R. (2008). 'Construction after Construction' and Its Theoretical Challenges. *Language*, 8-28.

Jackendoff, R. (2009). Compounding in the Parallel Architecture and Conceptual Semantics. In R. Lieber & P. Štekauer (Eds.) *The Oxford Handbook of Compounding* (105-128). Oxford: Oxford University Press.

Jolly, J. (1991). *Prepositional analysis within the Framework of Role and Reference Grammar*. New York: Peter Lang.

Katunar, D. (2015). *Ustroj leksikona u konstrukcijskoj gramatici – primjer prijedloga u hrvatskom jeziku.* Unpublished PhD thesis. Zagreb: Filozofski fakultet.

Katunar, D. & Raffaelli, I. (2018). Contrastive study of governed prepositions in Croatian, English and French. In P. Biały & B. Cetnarowska (Eds.) *Studies in Contrastive Semantics, Pragmatics and Morphology* (55-70). Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Kerovec, B. (2012). *Kognitivnolingvistička analiza prostornih odnosa u turskom i hrvatskom jeziku*. Unpublished PhD thesis. Zagreb: Filozofski fakultet.

Lyons, J. (1995). *Linguistic Semantics: An introduction.* Cambridge: Cambridge University Press.

Makkai, A. (1972). *Idiom structure in English* (Vol. 48). The Hague: Walter de Gruyter.

Martinet, A. (1966). Le mot. *Problèmes du langage*. Paris: Gallimard. 39-53.

Martinet, A. (1989). *Fonction et dynamique des langues.* Paris: Armand Colin.

Paillard, D. (2002). Prépositions et rection verbale. *Travaux de linguistique*(1), 51-67.

Petrović, V. (2016). Imenske skupine s prijedložnim komplementima u njemačkom i njihovi ekvivalenti u hrvatskom jeziku. *Jezikoslovlje* 17(1-2), 207-234.

Pranjković, I. (2002). *Hrvatska skladnja: Rasprave iz sintakse hrvatskoga standardnog jezika, drugo, izmijenjeno izdanje.* Zagreb: Hrvatska sveučilišna naklada.

Raffaelli, I. (2015). *O značenju. Uvod u semantiku.* Zagreb: Matica hrvatska.

Raffaelli, I. & Kerovec B. (2017). The concept of 'taste' in formation of Croatian and Turkish lexicon: A contrastive analysis. *Suvremena lingvistika*(83). 21-48

Rauh, G. (1993). On the grammar of lexical and non-lexical prepositions in English. In C. Zelinsky Wibbelt (Ed.). *The Semantics of Prepositions: From mental processing to Natural Language Processing* (99 – 151). Berlin, New York: Mouton de Gruyter.

Silić, J. & Pranjković, I. (2007). *Gramatika hrvatskog jezika za gimnazije i visoka sveučilišta*. Zagreb: Školska knjiga.

Tafra, B. (2005). *Od riječi do rječnika.* Zagreb: Školska knjiga.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language typology and syntactic description*, *3*(99), 36-149.

Dictionaries

Filipović, R. (2004). *Englesko-hrvatski rječnik.* Zagreb: Školska knjiga.

Hornby, Albert Sydney, & Sally Wehmeier (1995). *Oxford advanced learner's dictionary*. (Vol. 1428). Oxford: Oxford University Press.

LEXILOGOS: http://www.lexilogos.com/english/french_dictionary.html

Šonje, Jure. (2000). *Rječnik hrvatskoga jezika*. Zagreb: Leksikografski zavod Miroslav Krleža i Školska knjiga.

Hrvatski jezični portal, Novi Liber, SRCE, http://hjp.znanje.hr/

LingBaW

Linguistics Beyond And Within 4 (2018), 92-107

Patterns of phonosemantic reduplication in Kartvelian (South Caucasian) languages*

Zaal Kikvidze, Rusudan Gersamia, Maia Lomia
Tbilisi State University, Georgia

Abstract

In terms of phonosemantic doubling, root reduplication (in combination with affixation) is the most productive technique in Kartvelian languages (Georgian, Megrelian, Laz, Svan). The paper is a description of patterns of Kartvelian phonosemantic reduplication with respect to their both morphological and phonological parameters. The following types have been identified: 1. Root reduplication; 2. Syllable reduplication; 3. Redupluication with affixation. Each type has its respective sub-types. Based on abounding empirical data, the paper is an attempt to scrutinize and detect whether and how the above mentioned patterns are valid for all the four Kartvelian languages and to draw inferences about occurring formal and/or functional regularities associated with phonosemantic reduplication.

Keywords: phonosemantic reduplication, Kartvelian languages, reduplication patterns

1. Introduction

Since the present paper is an attempt to establish principal patterns of reduplication, specifically, those of phonosemantic reduplication, in Kartvelian languages, initially we will provide explicit definitions for the key issues and notions to be dealt with. Kartvelian languages constitute a branch of the Caucasian language family, and, since all of them are spoken in the southern part of the Caucasus, they are otherwise referred to as South Caucasian. The ISO 639-5 code for the branch is *ccs*; Glottolog: *kart1248*. As for the individual languages, they are the following: Georgian (ISO 639-3: *geo/kat*; Glottolog: *nuc11302*), Svan (ISO 639-3: *sva*; Glottolog: *svan1243*); Megrelian (ISO 639-3: *xmf*; Glottolog: *imng1252*), Laz (ISO 639-3: *lzz*; Glottolog:

* The present paper is part of the project *Phonosemantic Vocabulary in Kartvelian Languages (Structural, Semantic and Pragmatic Aspects)*, being implemented by the financial support from Shota Rustaveli National Science Foundation (Grant No 31/20). Any idea in this text is possessed by the authors and may not represent the opinion of Shota Rustaveli National Science Foundation.

lazz1240); many linguists conceive of Megrelian and Laz as dialects of the language called Zan (ISO 639-3: no code; Glottolog: *zann1245*).

> 'According to the standard interpretation, Mingrelian and Laz, which once formed a dialect continuum, are genetically closer to Georgian, with the archaic Svan language standing somewhat apart from its sister languages. There are, however, features that cut across genetic boundaries, for instance the development of additional evidential forms found in Mingrelian, Laz, Svan and some western and southern Georgian dialects. Similarly, specific analytic future tense formations (derived from 'to want') are found in some dialects both of Laz and of Southern Georgian, etc. In general, these languages have been in contact with each other for a very long time, and many features must have spread across dialect and language boundaries.'
>
> (Boeder 2005: 6-7)

Kartvelian languages have a rich consonant system where stops and affricates are distinguished according to a triadic opposition between voiced, voiceless aspirate, and ejective. For all of their differences, Kartvelian languages have preserved fundamentally the same agglutinative verb structure. Their peculiar structural properties include split ergativity and verb polypersonalism; verbs can show agreement with the subject, object, and indirect object. The verbal system is very complex; the languages make two distinctions, between stative or active and transitive or intransitive verbs. Tense/Aspect/Mood divides into series, and the forms of subject, object, and indirect object agreement marked on the verb vary by series.

All the Kartvelian languages are rich in dialect varieties:

> 'Identification of dialects of the Kartvelian languages is based on the linguistic (complex) principle – in characterizing the dialects, phonetic, morphological, syntactic and lexical peculiarities of each dialect are taken into consideration.'
>
> (Jorbenadze 1991: 19)

As for reduplication, since the time when the earliest classification of the phenomenon appeared back in 1862, introduced by August Friedrich Pott dividing *doubling* ('Doppelung') into two sub-types: *gemination* and *reduplication*, and conceiving of gemination, as different from the present-day understanding of the term, as *total doubling* ('Wiederholung im Ganzen') and of reduplication as *partial doubling* ('verkürzte und nur zum Theil, also bloß andeutungsweise vollzogene Wiederholung') (Pott 1862: 16), no drastic changes have been initiated in literature (for instance, Hurch 2005; Inkelas & Zoll 2005; Moravcsik 1978; Raimy 2000; Stolz et al. 2011). Hence, in the present paper we accept the following definition: reduplication is a phono-morpho-syntactic process whereby morphological and phonological items are repeated either totally or partially or with a slight modification. Thus, its inventories are twofold: phonological – syllables and moras, and morphological – roots, stems, affixes. As for phonosemantics, another key notion to be dealt with in the paper, it is viewed as an agglomeration of onomatopoeia and sound symbolism; hence, phonosemantic reduplication is a specific kind of the phenomenon in question manifesting some correlation between sound and meaning. Some linguists prefer the term 'expressives' to refer to the same phenomena:

> 'Expressives constitute a special form class (or part of speech) which differs in many ways from other, more familiar form classes such as nouns, verbs, or adjectives. Most commonly, expressives are very restricted syntactically (occurring in isolation or after one of a small number of words).'
>
> (Holisky 1988: 53)

The most noteworthy fact about expressives is that they do not always observe the phonotactic patterns established for the rest of the vocabulary (e.g., a stop which never occurs initially in other words may occur initially in expressives); frequently, expressives have their own phonotactic regularities (e.g. frequently, reduplicated shape) (Emeneau 1980_1: 264). There are a number of terms pertaining to non-arbitrary relationships between sound shape and meaning in language (see, for instance, Kikvidze 2014). Judging from the aforementioned, there is a need to clearly define and distinguish these and related terms and notions; linguists, referring the term 'expressives', offer some terminological distinctions. For instance, following Diffloth (1976: 263-264), Murray Emeneau states that

> ''expressives' is the most inclusive term for a form class with semantic symbolism and distinct morphosyntactic properties; 'ideophones' are a subclass in which the symbolism is phonological; 'onomatopoetics' are ideophones in which the reference of the symbolism is acoustic (i.e. imitative of sounds).'
>
> (Emeneau 1980_2: 7)

Following Voronin (1982) and Magnus (2001), we choose 'phonosemantic' and our choice is due to the fact that it incorporates all kinds of sound-meaning causal relationships.[1] However, we will maintain the notion 'expressives' as far as it can be a suitable reference for the specific cases of reduplication we are going to discuss, and, moreover, with respect to minor changes in vowels and/or consonants correlating with minor changes in meaning with particular emphasis on the fact that

> '[t]hese variations are often quite systematic, occurring across many sets of words and showing clearly that expressives are not subject to the condition of "lexical discreteness". Rather, among expressive it is often the case that incremental changes in the shape of a word (e.g., from high to mid to low vowel or from voiced to voiceless consonant) give corresponding incremental changes in the meaning (from small to medium to large, from less intense to more intense, etc.).'
>
> (Holisky 1988: 55)

Another very important point which we want to emphasize is that the focus on phonosemantics logically excludes non-phonosemantic reduplication from our scope.[2]

As it was already noted the aim of the present paper is to establish principal patterns of phonosemantic reduplication in Kartvelian languages. For the sake of the achievement of the goal, we identify three principal types (Syllable Reduplication, Root Reduplication, and Reduplication with Affixation), and, sometimes, relevant sub-types. Hence, the discussion is organized in line with the said structural patterns.

---

[1] We acknowledge and appreciate the fact that it was Stanislav Voronin (1982) who introduced the said understanding of phonosemantics; however, it should also be noted that the term itself (*phono-sémantique*) appeared much earlier (Pelliot 1936).

[2] Hence, we will not consider cases of distributive reduplication (Gil 1988; Choe 1991).

2. Types of reduplication in Kartvelian languages

Reduplication in Kartvelian languages has been dealt with by a number of scholars; the most notable of them are the following: Sanikidze (1968; 1976; 1977), Topuria (1979), Datukishvili (1990), Aronia (2010), Gersamia (2015), etc. Hence, various linguists have identified various patterns of Kartvelian reduplication, such as: stem reduplication, distinguishing between modified and non-modified stems (Shanidze 1955; Sanikidze 1976; Topuria 1979); syllable reduplication, distinguishing between open and closed syllable patterns (Neisser 1953; Zhgenti 1960); total and partial reduplication (Datukishvili 1990); split and non-split reduplication, distinguished in accordance with the possibility of insertion of a linking element between a base and a reduplicant (Datukishvili 1990).

As far as the present paper is a description of Kartvelian phonosemantic reduplication patterns in terms of their both phonological and morphological parameters and with a view to structural features of the language in point, we prefer to identify the following types:

2.1. Syllable Reduplication

2.2. Root Reduplication

2.3. Reduplication with Affixation

Each of them may have several sub-types. Therefore, we organize our discussion according to the identified types and sub-types.

2.1. Syllable reduplication

Two kinds of syllables are doubled in most of the Kartvelian languages; hence, there are two sub-types of syllable reduplication:

2.1.1. Open syllable: (CV)[2]

Georgian
***bi**: bi-bi-n-i* 'waving, swaying (grass)'[3],
***pi:** pi-pi-n-i* 'twittering, rustling'
***k'a:** k'a-k'a-n-i* '(hen's, pheasant's) cackling; cackling speech; (machine-gun's) rattle'
***ka:** ka-ka-n-i* 'panting; endless talk; ranting'
***t'a:** t'a-t'a-n-i* 'scolding, remonstrations'

Megrelian
***k'i:** k'i-k'i-n-i* 'gazing'
***zi:** zi-zi-n-i* 'swollen; bloated (with food)'
***ži:** ži-ži-n-i* 'to get soaked'
***ği:** ği-ği-n-i* 'kindling; flaring (fire)'
***ši:** ši-ši-n-i* 'eating one's fill; being sated'

Laz
***č'e:** č'e-č'e* 'chatterbox, gossip'
***ča:** ča-ča-ra* 'blethering'
***ğu:** o-ğu-ğu-l-u* 'cooing; smoldering'
***ʒi:** go-ʒi-ʒi-l-u* 'loud laughter'

Instances of open syllable doubling have not been evidenced in any of the dialects of Svan.

[3] All translations into English are provided in accordance with Rayfield (2006) if not otherwise stated.

2.1.2. Closed syllable: (CVC)[2]

Georgian	***bax**: bax-bax-i* 'bang; bragging'
	***rak':** rak'-rak'-i* '(water) glugging; (stream) burbling'
	***čip:** čip-čip-i* 'mumbling/gabbling toothlessly'
	***k'is:** k'is-k'is-i* 'peels of laughter'
	***ğav:** ğav-ğav-i* 'yapping; steady barking'
Megrelian	***xar:** xar-xal-i*[4] 'loud laughter, guffaws'
	***t'an:** t'an-t'al-i* 'chattering, pratling'
	***var:** var-val-i* 'heating (metal) white-hot, red-hot'
	***c'ir:** c'ir-c'in-i* '(child's) cry'
	***zur:** zur-zul-i* 'wailing, keening'
	***par:** par-pal-i* 'rocking (of boat on sea), swaying (of grass in wind)'
Laz	***t'ir:** t'ir-t'ir-i* 'chattering, prattling'
	***par:** par-pal-i* 'butterfly'
	***gur:** gur-gul-i* 'thunder'
	***bar:** bar-bal-i* 'silly talk'
	***dar:** dar-dal-i* 'silly behaviour'
Svan	***ǯul:** ǯul-ǯun-Ø* '(animal's) jog-trot; (sb's) lumbering run'
	***ğur:** ğur-ğun-Ø* 'roar (of rushing water); loud bubbling (of boiling water)'
	***pər:** pər-pən-Ø* 'silly talk'
	***bitk:** bitk-bitk-Ø* 'banging'
	***čix:** č'ix-č'ix-Ø* 'crackling (of gunfire)'

2.2. Root reduplication

In terms of phonosemantic doubling, root reduplication (rather frequently, in combination with affixation) is the most productive technique in Kartvelian languages. With respect to the common pattern of the root structure in Kartvelian languages, two sub-types are identified:

2.2.1. Sub-type (CVC)[2]

Georgian	***bax**: bax-bax-i* 'bang; bragging'
Svan	***ğur:** ğur-ğur* 'roar (of rushing water); loud bubbling (of boiling water)'
Megrelian	***xar:** xar-xal-i* 'loud laughter, guffaws'
Laz	***t'ir:** t'ir-t'ir-i* 'chattering, prattling'

2.2.2. Sub-type (CV)[2]

Georgian	***bi**: bi-bi-n-i* 'waving, swaying (grass)'
Megrelian	***ži:** ži-ži-n-i* 'to get soaked'
Laz	***č'e:** č'e-č'e* 'chatterbox'

In both cases, consonant clusters may develop and increase both in anlaut and auslaut positions:

[4] In the reduplications like ***xar:** xar-xal-i*, the dissimilation process (**xar-xar-i* > *xar-xal-i*) takes place. This is a homogeneous process for all the Kartvelian languages.

Root-initial cluster $(C_NVC)^2$

Georgian	***brax:*** *brax-brax-i* 'loud stamping'
Megrelian	***ǯǧir:*** *ǯǧir-ǯǧin-i* 'speaking nasally'
Svan	***č'q'wip':*** *č'q'wip'-č'q'wip'-Ø* 'snapping (thread)'

Root-final cluster $(CV_N)^2$

Georgian	***batk:*** *batk-a-butk-i* '(guns) banging'
Svan	***bitk:*** *bitk-bitk-Ø* '(guns) banging'

In Georgian and Svan, sub-type $(CVC)^2$ has two realizations: Total and Partial (Ablaut). In case of total reduplication, the exact representations of a base and a reduplicant are the following:

2.2.1.(T) CVC+CVC—NOM

Georgian	***bax****: bax-bax-i* 'bang; bragging'
	rak': *rak'-rak'-i* '(water) glugging; (stream) burbling'
Svan	***bitk****: bitk-bitk* '(guns) banging'
	šk'wip: *šk'wip-šk'wip* '(noise of) beating/thrashing sb/sth (with a stick)'

In case of ablaut republication, they are:

2.2.1.(A) CV_1C+LINK+ CV_2C—NOM

Georgian	***bax****:*	*bax-a-bux-i* 'swagger'
		bax-i-bux-i 'swagger'
	rak':	*rak'-a-ruk'-i* '(woodpecker) drumming'
		rak'-i-ruk'-i '(woodpecker) drumming'
Svan	***bitk****:*	*bitk-i-batk(a)* '(guns) banging'
	šk'wip:	*šk'wip-i-šk'wip(a)* '(noise of) beating/thrashing sb/sth (with a stick)'

Judging from the above presented formulae, some patterns of syllable and root reduplication seem to coincide. In fact, in terms of their forms, the formulae for 2.1.1. and 2.2.1. do coincide: $(CVC)^2$ (Kikvidze 2018).

However, that does not imply that they are one and the same. Functionally, the bases are different (Kikvidze 2018); in 2.1.1., it is a morphological entity, that is a root, while, in 2.2.1., it is a prosodic entity, that is a syllable. Structurally, the difference is realized in the fact that, in case of 2.2.1., a base, as a root, is both doubled (2.2.1.(T) CVC+CVC—NOM and reduplicated through vowel alternation (2.2.1.(A) CV_1C+LINK+CV_2C—NOM), while, in case of 2.1.2., a base, as a syllable, can only be doubled (copied) (CVC+CVC—NOM).

3. Reduplication with affixation

This pattern (Reduplication with Affixation) has been identified due to various reasons: 1) it is a very productive phenomenon; 2) affixes are an integral part of the reduplication process in Kartvelian languages. Three kinds of affixation processes have been noted: suffixation, circumfixation, and interfixation. We will describe each of the sub-types in turn.

3.1. Reduplication with suffixation

The formula for the suffixation pattern is the following: CV-CV+SUF—NOM.

The pattern in point has a very clear-cut feature: only open syllables occur within this formation. As for consonants, they can be all the stops and affricates (voiced, voiceless aspirate, and voiceless ejective) and fricatives (voiced and voiceless); sonorants (*m, n, r, l*) do not appear in such formations. Of the five vowels, only three of them (*a, e, i*) occur. Various scholars have observed regular correlations of consonantal features to particular shades of meaning. For instance, N. Tschanischwili (1988: 171) notes that voiced consonants correlate with 'strong, high pitch', while voiceless aspirated stops and voiceless fricatives are associated with the meanings 'weak, gentle, sensitive', and ejectives correlate with salience, stridency. With respect to these correlations, she identifies several groups of synonyms of phonosemantic reduplicative verbs:

1) verbs referring to noise: strong (*bu-bu-n-eb-s* 'lowing, bellowing (by bull, stag, etc.)', *du-du-n-eb-s* 'murmur, mumbling', *gu-gu-n-eb-s* 'low roar, engine sound; crackling (fire)', *zu-zu-n-eb-s* 'swish, whistle (e.g. of whip, bullets); roar (of wind); rustling (of leaves), buzzing (of bees)' and weak (*si-si-n-eb-s* '(goose, wind) hissing; shushing (noisy child); saying "psst" (to get sb's attention)', *ši-ši-n-eb-s* 'soft bubbling (of sth boiling); fizzling (of damp firewood); spluttering', *xi-xi-n-eb-s* 'wheezing');
2) verbs referring to sounding, speaking: neutral (*ka-ka-n-eb-s* '(hen's, pheasant's) cackling; cackling speech; (machine-gun's) rattle)', weak (*t'i-t'i-n-eb-s* 'burbling; drivel; twaddle', *č'i- č'i-n-i* 'whirring, rattling (of cricket, etc.); whine (of bag-pipes); fizz (of fermenting wine)', and strident (*c'u-c'u-n-eb-s* 'lamenting, wailing; (mosquito) whine', *t'a-t'a-n-eb-s* 'scolding, remonstrations', *c'i-c'i-n-eb-s* '(bird's) cheeping', *k'i-k'i-n-eb-s* 'bleating (goat)');
3) verbs referring to singing: weak (*ğı-ğı-n-eb-s* 'singing quietly', *ğu-ğu-n-eb-s* 'cooing; smouldering');
4) verbs referring to motion: weak (*bi-bi-n-eb-s* 'waving, swaying (grass)', *pi-pi-n-eb-s* 'twittering, rustling') and strident (*p'i-p'i-n-i* '(bag-pipe, chalumeau's) piping (noise)', *č'i- č'i-n-i* 'whirring, rattling (of cricket, etc.); whine (of bag-pipes); fizz (of fermenting wine)');
5) characteristic feature: weak (*(ga)-p'i-p'i-n-eb-ul-i* 'full to the brim', *ci-ci-n-i* 'whirring, rattling (of cricket, etc.); whine (of bag-pipes); fizz (of fermenting wine)', strong (*(a)žu-žu-n-eb-s* 'make eyes (at sb)' (ibid.: 175).

As for the vowels, ***a*** is associated with neutral meaning (*ka-ka-n-eb-s* '(hen's, pheasant's) cackling; cackling speech; (machine-gun's) rattle)', *ba-ba-n-eb-s* 'shiver', *q'a-q'a-n-eb-s* 'jabbering, gaggling; croaking', *č'a- č'a-n-eb-s* 'the slightest trace', *t'a-t'a-n-eb-s* 'scolding, remonstrations'); in combination with voiced consonants, ***i*** correlates with the meaning 'little, weak' (*si-si-n-eb-s* '(goose, wind) hissing; shushing (noisy child); saying "psst" (to get sb's attention)', *t'i-t'i-n-eb-s* 'burbling; drivel; twaddle', *ši-ši-n-eb-s* 'soft bubbling (of sth boiling); fizzling (of damp firewood); spluttering', *c'i-c'i-n-eb-s* '(bird's) cheeping'); in combination with voiced consonants, ***u*** correlates with the meaning 'large, strong' (*gu-gu-n-eb-s* 'low roar, engine

sound; crackling (fire)', *bu-bu-n-eb-s* 'lowing, bellowing (by bull, stag, etc.)', *zu-zu-n-eb-s* 'swish, wistle (e.g. of whip, bullets); roar (of wind); rustling (of leaves), buzzing (of bees)') (ibid.: 175).

In the similar way, based on expressive manner-of-speak verbs in Georgian, Holisky (1988) and Holisky & Kakhadze (1988) provide the sound symbolic associations of the phonemes of Georgian: (a) Vowels: /u/ – softer, lower pitch; /i/ – high pitch, louder; /a/ and /o/ – loud, more negative (When paired with similar verb with different vowel);[5] (b) Phonation type: glottalized – higher pitch, more positive; voiceless and voiced – lower pitch, more negative; (c) Place of articulation: bilabials – soft, low pitch; velar versus uvular stops – higher versus lower pitch; velar spirants – louder, more negative (Holisky & Kakhadze 1988: 198-190). This is true for reduplicated manner-of-speaking verbs and nomina actionis in their sample. We picked reduplicatives from their list in which the authors observed their meanings (and shades of meaning):

k'is-k'is-i it belongs to the pattern $(CVC)^2$; hence, it takes no affixes. Here is how they authors describe its meaning: 'laugh, of a young girl in good mood, with high pitch, silvery voice, somewhat loud, pleasant' (ibid.: 195);

rox-rox-i $(CVC)^2$: 'emit loud, bumpy irregular noise, of man with loud, base, resonant voice, often talking and laughing simultaneously' (ibid.: 197);

q'a-q'a-n-i it belongs to the pattern $(CV)^2$; hence, it takes on the derivational suffix -***n***: 'emit loud noises of many voices simultaneously, perhaps of a boisterous crowd, slightly derogatory (in that it implies the speakers are being too loud and impolite)' (ibid.: 197);

t'u-t'u-n-i $(CV)^2$: '(colloquial) converse very softly, sweetly and indistinctly' (ibid.: 198);

k'ur-k'ul-i $(CVC)^2$: '(colloquial) coo or flirt quietly, of man and woman to each other (like two doves)' (ibid.);

but'-but'-i $(CVC)^2$: 'mumble softly, not particularly articulately, often to oneself' (ibid.);

du-du-n-i $(CV)^2$: 'talk quietly, indistinctly, often of groups of people; noise of moving water, e.g., a wide slow river' (ibid.);

čur-čul-i $(CVC)^2$: 'whisper' (ibid.);

ču-ču-n-i $(CV)^2$: 'talk very softly to oneself, perhaps semi-consciously, while engages in a chore or light activity (like dusting or puttering around)' (ibid.);

si-si-n-i $(CV)^2$: 'hiss quietly to someone; subject is often angry and insulting or reprimanding the indirect object; expresses very negative feelings about subject' (ibid.: 198);

zu-zu-n-i $(CV)^2$: 'emit a [zzzz] sound, of wind, insects, bomb' (ibid.);

čip-čip-i $(CVC)^2$: 'talk, of toothless old person (or someone imitating one)' (ibid.: 199);

luq'-luq'-i $(CVC)^2$: 'talk incoherently due to some physical cause (exhausted, drunk, brain damaged from stroke) or because one is extremely unnerved and can't get words out' (ibid.);

bu-bu-n-i $(CV)^2$: 'talk continuously in a monotone, low pitch voice' (ibid.: 200);

[5] It is noteworthy that "/e/ never participates in symbolism (it is not found in a single manner of speaking verb), and /o/, which is very rare, echoes /a/" (Holisky & Kakhadze 1988: 191).

ǯiq'-ǯiq'-i (CVC)²: '"talk incoherently; sound is very unpleasant, almost animal-like, as though subject is trying to talk while about to throw up; conveys speaker's extremely negative evaluation of the speech' (ibid.);

viš-viš-i (CVC)²: 'lament or moan (necessarily out loud, rather loudly and showing aggravation) over something which has gone wrong or which may go wrong' (ibid.: 201);

k'uč'-k'uč'-i (CVC)²: 'laugh, of women, children, young men, softly, with high pitch, endearing' (ibid.: 202);

xit-xit-i (CVC)²: 'laugh, with lower pitch, softer, perhaps choking a bit with laughter' (ibid.);

xar-xar-i (CVC)²: 'laugh loudly, mostly of larger people, perhaps body is shaking with laughter (pitch is more neutral)' (ibid.: 203);

xvi-xvi-n-i (CV)²: 'laugh, of a man in ugly, unpleasant manner, implies that he lacks social graces' (ibid.);

xi-xi-n-i (CV)²: 'emits rasping sound […], perhaps during heavy breathing' (ibid.);

q'ič'-q'ič-i (CVC)²: 'higher pitch rasping, caused by a vibration of a flabby substance, e.g. flesh in throat of a fat person' (ibid.: 204);

ka-ka-n-i (CV)²: 'non-stop talking of a woman while walking around, waving her arms, giving orders' (ibid.);

č'ik'-č'ik'-i (CVC)²: 'animal: emit noise, of a swallow; human: talk, of small child, with lively, sweet, crisp voice, speaking in excited or happy manner; speaker has pleasant feeling toward subject' (ibid.: 205);

t'ik'-t'ik'-i (CVC)²: 'talk actively, of a small child, with crisp, sharp voice' (ibid.);

t'i-t'i-n-i (CV)²: 'talk, of small child, with childish, naïve, sweet, soft voice' (ibid.);

ğu-ğu-n-i (CV)²: 'animal: coo, of a dove; human: emit [aaa] and [uuu], of small baby; nut talking but sweet cooing' (ibid.).

The discussed data shed light on the structural dimensions of the pattern in point. It is particularly significant to have established that suffixation is only possible whenever an open syllable, as a phonosemantic root, is reduplicated: (CVC)²; specifically, CV-CV-*n*[6]-*i*. Besides, no derivational suffix can be taken on by a closed syllable as a phonosemantic root; hence, *CVC-CVC+SUF—NOM is not possible.

3.2. Reduplication with circumfixation

The formula for the circumfixation pattern is the following: INF<CVC>INF

Laz is the only Kartvelian language to exemplify reduplication with circumfixation (o- -u):

[6] The suffix in question (-***n***) does not appear in the *Dictionary of Morphemes and Modal Elements of the Georgian Language* (Jorbanedze et al. 1988).

Table 1: *Laz circumfixated reduplication*

Simplex	Reduplicative	Translation
t'ar	*o-t'ar-t'al-u*	'to chatter'
č'ir	*o-č'ir- č'il-u*	'lamenting, wailing; (insect) whine'
kir	*o-kir-kin-u*	'peels of laughter'
xar	*o-xar-xal-u*	'loud laughter, guffaws'
dar	*o-dar-dal-u*	'silly talk'

It should necessarily be noted that circumfixation is not an exclusive mechanism for phonosemantic reduplication in Laz; by means of the circumfixated formation, the language strives to fit phonosemantic reduplication into the pattern characteristic of regular masdars.

Irrespective of the fact that this sub-pattern (Reduplication with Circumfixation) occurs only in Laz, structurally it is quite typical of the Kartvelian languages at large.

3.3. Reduplication with interfixation

The formula for this pattern (Reduplication with Interfixation) is the following: CV_1C+LINK+CV_2C—NOM

Of all the Kartvelian languages, Georgian has the most clear-cut system of interfixated phonosemantic reduplication. It should be noted that this pattern is necessarily accompanied by vowel alternation in reduplicants as it is reflected in the formula. More specifically, the alternation is as follows: ***a-u***. Hence, the formula can be specified as 3.3.1. C***a***C+***a/i***+C***u***C—NOM. It was the Georgian linguist Parnaoz Ertelishvili who identified them as an individual type of reduplicatives in Georgian:

> 'The stems are an outcome of root reduplication; in an initial syllable of the root, there is the vowel ***a***, while, in the final one, there is the ***u***. simple root are connected by means of the interfixal vowels ***a*** and ***i***; the ***a*** is more frequent. Parallel variants of stems, with respect to interfixal vowels, have been attested.'
>
> (Ertelishvili 1978: 70)

Individual varieties of the above presented formula are the following:

3.3.1. CV_1C+LINK+CV_2C—NOM, specifically C***a***C+***a/i***+C***u***C—NOM

bax: *bax-a-bux-i* 'bang; bragging'

3.3.2. CCV_1C+LINK+CCV_2C—NOM, specifically CC***a***C+***a/i***+CC***u***C—NOM

čxar: *čxar-a-čxur-i* 'loud constant jangling (metal, glasses)'

3.3.3. $CCCV_1C$+LINK+$CCCV_2C$—NOM, specifically CCC***a***C+***a/i***+CCC***u***C—NOM

txlaš: *txlaš-a-txluš-i* 'a series of slaps; slopping (of feet in mud); stamping, clopping (of feet, hooves)'

3.3.4. CV_1CC+LINK+CV_2CC—NOM, specifically C***a***CC+***a/i***+C***u***CC—NOM

rac'k': rac'k'-a- ruc'k' 'tinkling, jingling'

He estimates up to 120 interfixal reduplicatives in Georgian (Ertelishvili 1970: 78).

With respect to the two vowel interfixes, illustrations of parallel variants are the following:

Table 2*: Georgian interfixated reduplication*

Simplex	Interfix -***a***	Interfix –***i***	Translation
t'ak'	*t'ak'-a-t'uk'-i*	*t'ak'-i-t'uk'-i*	'tick-tock, tick-tocking'
č'ax	*č'ax-a-č'ux-i*	*č'ax-i-č'ux-i*	'crackling (of gunfire)'
č'a č'q'	*č'a č'q'-a-č'u č'q'-i*	*č'a č'q'-i-č'u č'q'-i*	'squelching (of feet in bog)'
p'ak'	*p'ak'-a-p'uk'-i*	*p'ak'-i-p'uk'-i*	'patter, clack, clatter (of feet, heels, hooves)'
sxap'	*sxap'-a-sxup'-i*	*sxap'-i-sxup'-i*	'gabbling'
partx	*partx-a-purtx-i*	*partx-i-purtx-i*	'fluttering, flapping (of wings)'

Concerning such reduplicatives, there is a distinct approach according to which "interfixated compounding does not occur in Georgian" (Datukishvili 1990: 44). She views the above mentioned reduplicatives as coordinating compounds the components of which (a base and a reduplicant) can be conjoined by means of the conjunction *da* 'and' (ibid.: 42). Therefore, not roots but rather stems are reduplicated. Hence, the ***a*** and ***i***, having been presented as interfixal vowels, are conceived as an integral part of a base (***a***: *baxa-*, *t'a'k'a-*, *č'axa-*, *č'a č'q'a-*, *rak'a-*, *p'aka-*, *sxap'a-*, *partxa-*; ***i:*** *baxi-*, *t'a'k'i-*, *č'axi-*, *č'a č'q'i-*, *rak'i-*, *p'aki-*, *sxap'i-*, *partxi-*). Thus, the reduplicatives will be glossed not as it was presented in Table 1, but in a rather different way, as it is in the following: ***bax***: *baxa-bux- // baxi-bux-*; ***t'ak'***: *t'ak'a-t'uk'- // t'ak'i-t'uk'-*; ***č'ax:*** *č'axa-č'ux- // č'axi-č'ux-*; ***č'a č'q':*** *č'a č'q'a-č'u č'q'- // č'a č'q'i-č'u č'q'-*; ***rak':*** *rak'a-ruk'- // rak'i-ruk'-*; ***p'ak':*** *p'ak'a-p'uk'- // p'ak'i-p'uk'-*; ***sxap':*** *sxap'a-sxup'- // sxap'i-sxup'-*; ***partx:*** *partxa-purtx- // partxi-purtx-*.

The author's argumentation is based upon the diachronic data:

> 'Onomatopoeic syllables took on the word-formation suffix -***a***, resulting in nouns for noise: *lac'- > *lac'-a, brax- > *brax-a, t'k'ac- > *t'k'ac-a*, which later were reduplicated… Reduplication and vowel alternation should have occurred simultaneously.'
>
> (ibid.: 44)

The contradicting approaches can be represented in the following way:

Table 3: *Interfixated or not?*

	Variant with the Vowel –***i***	Variant with the Vowel –***a***
(1)	batk-i-butk-i	batk-a-butk-i
	batk-SG.NOM-butk-SG.NOM	batk-DER-butk-SG.NOM
(2)	batk-i-butk-i	batk-a-butk-i
	batk-LINK-butk-SG.NOM	batk-LINK-butk-SG.NOM
(3)	batki-butk-i	batka-butk-i
	batki-butk-SG.NOM	batka-butk-SG.NOM

With respect to the aformentioned, the principal question to be asked is about the status of the base: is it a stem or a root? If it is conceived of as a root (Ertelishvili 1970), the inter-radical vowels ***a*** and ***i*** are interfixes; if it is considered to be a stem (Datukishvili 1990), then the vowels in question are not interfixes but rather are assumed as an integral part of a stem structure.

Normally, such questions can be more easily answered with respect to data from sister languages. Therefore, we have to see what the situation is in Svan, Megrelian and Laz in order to find out whether similar patterns occur in them, and how the attested patterns correlate in terms of their structural and semantic properties.

Varlam Topuria (1979: 118) names only two examples for Megrelian: *k'iru-k'aru* 'ritual cake' and *zik'u-zak'u* 'a swing'. Later, some more illustrations were drawn in the scholarly literature: *čim[u]- čamu* 'tiny noise', *piču-paču* 'sluggish', *t'iru-t'aru* 'to chatter', *p'int'i-p'ant'u* 'various tiny, small things' (Aronia 2010: 42), *xiru-xaru* 'heap of junk, old rubbish' (Kiria et al. 2015: 225).

Similarly to Megrelian, Laz data evidence only a few examples: *č'ak'a-c'uk'a* 'of all colours', *bric'u-brac'u* 'ripping (sound)', *mč'ipe-mč'upe* '(a lot of) tiny, small things' (Aronia 2010: 44), *baga-buga* 'pounding, thumping (of a heart)', *ğala-ğula* 'untidy', *ğač'a- ğuč'u* 'everyone talking at a time', *c'ak'ara-c'uk'uru* 'anything at hand', *xvit'i-xvat'a* 'mottled' (Tandilava 2013: 46, 808-809, 850, 882).

Another Kartvelian language Svan seems to display an abounding number of reduplicatives of the pattern in question;[7] we are providing some of its typical illustrations in the following table:

Table 4: *Svan interfixated reduplication*

Simplex	Reduplicative	Translation
t'q'ig	*t'q'ig-i-t'q'aga*	'loitering, roaming about'
bitk	*bitk-i-bätka*	'(guns) banging'
bič'k'w	*bič'k'w-i-bäč'k'wa*	'(noise) of crackin'
p'ilt'	*p'ilt'-i-p'alt'a*	'(a child's or an adult thin woman's) unintelligible talk'
šk'wip	*šk'wip-i- šk'wapa*	'(noise of) beating/thrashing sb/sth (with a stick)'
riq'	*riq'-i-räq'a // riq'-i-raq'a*	'crashing, banging'

It should necessarily be noted that the Lashkh and Cholur varieties (sub-dialects) of Svan have the following forms: *bili-bulu* 'stammer', *biri-buru* 'silly talk', *p'iri-p'uru // k'ir-i-k'uru* 'silly, rapid, endless talk', *p'ic'i-p'uc'u* '(a pampered girl's) talk with changed, high pitch voice', *pirpi-purpu* 'slurring (speech defect)', *piči- puču* 'whisper', *picki-pucku* 'pampered girl'.[8]

The patterns of ablaut-motivated reduplication in Kartvelian languages are rendered in the following table:

[7] It is noteworthy that they appear in all of its dialects; this is a significant factor as far as dialects of Svan are not always uniform in terms of representation of various linguistic phenomena.

[8] All of the Svan examples were elicited within the framework of the above mentioned project and kindly provided by our colleague and collaborator within the project Dr. Medea Sagliani. She also dealt with the problem in question in one of her recently published papers (Sagliani 2015).

Table 5: *Ablaut-motivated reduplication in Kartvelian languages*

Georgian	*a-u*	*a-a-u-*i*	*batk-a-butk-i*	*k'nac'-a-k'nuc'-i*	*rak'-a-ruk'i*	*p'ak'-a-p'uk'i*	*čax-a-čux-i*
		*a-i-u-*i*	*batk-i-butk-i*	*k'nac'-a-k'nuc'-i*	*rak'-i-ruk'i*	*p'ak'-i-p'uk'i*	*čax-i-čux-i*
Megrelian	*i-a*	*i-u-a-u*	*zik'u-zak'u*	*k'iru-k'aru*	*t'iru-t'aru*	*p'it'[u]-p'at'u*	*piču-paču*
Laz	*a-u*	*a-a-u-a*	*č'ak'a-č'uk'a*	*k'rank-k'runk*			
	i-a	*i-u-a-u*	*bric'u-brac'u*				
Svan	*i-a*	*i-i-a/ä-a*	*bitk-i-b ätka*	*riq'-i-r äq'a*	*č'ipx-i- č'apxa*	*t'q'ig-i-t'q'aga*	*p'ilt-i-p'alt'a*
	i-u	*i-i-u-u*	*bir-i-buru*	*p'ir-i-p'uru*	*p'ic'-i-p'uc'u*	*p'ič-i-p'uču*	

As it is seen, vowel alternations are not the same in all the Kartvelian languages. Therefore, it can be questioned that Megrelian and Laz patterns have something in common with those of Georgian and Svan.

In Megrelian, both the base and the reduplicant end in -***u***. It should be noted that both -***u***'s are frequently truncated, and this can be explained in terms of the rule of narrowing of the vowels ***u*** and ***i***, which is characteristic of Megrelian: ***i/u>ə>Ø***: *čim-čamu < čimu-čamu, p'it' p'at'u < p'it'u-p'at'u, t'ir-t'aru < t'iru-t'aru*, etc. (Kartozia et al. 2010: 480). The segments (base and reduplicant) with the -***u*** refer to the immediate repetition of two completed actions with a changing direction of motion, that is, each constituent has an implication of finiteness, whereas, whenever they occur without the -***u***, and with either a schwa or a zero, the same meaning is rendered by reduplicatives: *čim-čam* (***u/ə/Ø***), *p'it'-p'at'* (***u/ə/Ø***), *t'ir-t'ar* (***u/ə/Ø***), etc. This specific pattern of Megrelian is similar to those in other Kartvelian languages (we mean reduplication with vowel alternation though without an interfix); cf.:

Svan: *biri-buru* 'silly talk'; *p'iri-p'uru // k'iri-k'uru* 'silly, rapid, endless talk'; *p'ic'i-p'uc'u* '(a pampered girl's) talk with changed, high pitch voice'; *pirpi-purpu* 'slurring (speech defect)'; *piči-puču* 'whisper'; *picki-pucku* 'pampered girl'

Georgian: *peri-puri* "'facial colour'; *k'aba-k'uba* 'oman's clothing, dressing'; *c'ağa-c'uğa* 'ankle-length boot/shoe'; *xara-xura* 'junk, old rubbish'; *parča-purča* 'silks'; *č'rela-črula* 'of all colours'; *č'ia-č'ua* 'worms, maggots'

Laz: *bric'u-brac'u* 'ripping (sound)'; *ğač'a-ğuč'u* 'everyone talking at a time'; *c'ak'ara-c'uk'uru* 'anything at hand'; *q'vili-q'vala* 'bone'; *xvit'i-xvat'a* 'mottled'; *kčini-kčvani*'old-old'; *mč'ipe-mč'upe* '(a lot of) tiny, small things'; *ğala- ğula* 'untidy'; *č'ak'a-č'uk'a* 'mottled'; *bara-bura* 'potato'

The Megrelian (k'iru-k'aru, zik'u-zak'u, čim[u]- čamu, piču-paču, t'iru-t'aru, p'int'i-p'ant'u, xiru-xaru) and Svan (biri-buru, p'iri-p'uru // k'iri-k'uru, p'ic'i-p'uc'u, pirpi-purpu, piči-puču, picki-pucku) phonosemantic reduplicatives belong to the pattern $(CVCV)^2$; specifically, Megrelian: $CV_1CV_2+CV_3CV_2$ = CiCuCaCu;[9] Svan: $CV_1CV_1+CV_2CV_2$ = CiCiCuCu. The rest of the items are not phonosemantic ones. The same can be stated about the Georgian list (peri-puri, k'aba-k'uba, c'ağa-c'uğa, xara-xura, parča-purča, č'rela-črula, č'ia-č'ua); they belong either to a noun or an adjective class, none of them being phonosemantic. The coda vowels are also different; the reduplicative-internal **i** and **u** should not be identified as functional entities. Hence, in the Megrelian examples, the internal **u/ə** should not be regarded to be an interfix similar to the Georgian **a** and **i** and the Svan **i**. Therefore, the occurrence of

[9] The Laz *bric'u-brac'u* corresponds to the same pattern.

phonosemantic reduplication with interfixation should be recognized only in Georgian and Svan. Their common structure can be represented by the following formula: CVC-V-CVC.[10] Hence, both a base and a reduplicant can only be closed syllables (CVC). Georgian interfixated reduplicatives take on a nominative case marker -**i** (CVC-V-CVC- **i**), while those of Svan have the -***Ø*** (CVC-V-CVC-Ø); the Svan final -**a** is attested only in poetic texts. It should also be noted that not all the **a**-interfixated reduplicatives have their **i**-interfixated counterparts; the latter are much fewer.

4. Conclusions

In the present paper, we attempted to scrutinize and detect whether and how the above mentioned patterns are valid for all the four Kartvelian languages and to draw inferences about occurring formal and/or functional regularities associated with phonosemantic reduplication.

It has been established that, with respect to the peculiarities in the languages in point, the most characteristic patterns of phonosemantic reduplication are Syllable Reduplication, Root Reduplication, and Reduplication with Affixation. Templatic similarities between some patterns of syllable reduplication and root reduplication were clarified highlighting structural and functional differences. In fact, the interplay of prosodic and morphological inventories is well demonstrated in the description of reduplication processes. In case of total $(CVC)^2$ reduplication, a template prosodic entity [CVC] is copied: [CVC-CVC]. As a result of the grammaticalization process, an onomatopoeic syllable turns into a root/stem. The resulting construction takes on a case marker: CVC-CVC-NOM (Georgian: CVC-CVC-***i***; Svan: CVC-CVC-***Ø***). Besides, the process is much more complicated in case of ablaut reduplication when a base is reduplicated through vowel alternation; a base and a reduplicant are conjoined by means of an interfix (***a*** or ***i***): *bax-a-bux-i // bax-i-bux-i*. More specifically, the process in point involves several stages: 1) a syllable template is fully reduplicated and linked to its C/V slots (C***a***C> C***a***C-C***a***C: ***bax***>*bax-bax*); 2) an ablaut-motivated vowel change occurs from /***a***/ to /***u***/ for the reduplicant (C***a***C-C***a***C> C***a***C-C***u***C: *bax-bax*>*bax-bux*); 3) interfixation as construction (compound) marking (C***a***C-LINK-C***u***C-NOM: *bax-a-bux-i* or *bax-i-bux-i*).

The described patterns of phonosemantic reduplication demonstrate common constraints of the grammaticalization/phonotactics interface in Kartvelian languages. With respect to both the inventory and the structural and functional features, it is obvious that phonosemantic reduplication in Kartvelian languages is a very clear-cut **morphonological** phenomenon.

References

Aronia, Inga. *Ko'omp'ozit'ebi megrul-lazurši [Compounds in Megrelian-Laz]*. Tbilisi: Universal, 2010.
Boeder, Winfried. South Caucasian languages. *Lingua* 115 (2005): 5-89.

[10] A base (CVC) of the CVC-V-CVC pattern is a productive element as a free form, taking on the verbal suffixes (-***un*** or -***an***); meanwhile finite verbs can never be derived from the reduplicatives themselves (CVC-V-CVC) (Kobalava 1980: 71).

Choe, Jae-Woong. Georgian reduplication and a relational theory of distributivity. In *The Sixth Japanese-Korean Joint Conference on Formal Linguistics*, edited by Akira Ikeya, 4-25. Tokyo: The Logico-Linguistic Society of Japan, 1991.

Datukishvili, Ketevan. *Redup'lik'acia kartulši [Reduplication in Georgian]*, PhD Dissertation. Tbilisi: Arnold Chikobava Institute of Linguistics, Georgian Academy of Sciences, 1990.

Diffloth, Gerard. Expressives in Semai. *Austroasiatic Studies (Oceanic Linguistics Special Publication 13)*, 249-264. Honolulu, Hawaii: University of Hawaii Press, 1976.

Emeneau, Murray B. Onomatopoetics in the Indian linguistic area. In *Language and Linguistic Area: Essays by Murray B. Emeneau, Selected and Introduced by Anwar S. Dil*, 250-293. Stanford: Stanford University Press, 1980_1.

Emeneau, Murray B. Linguistic area: Introduction and continuation. In *Language and Linguistic Area: Essays by Murray B. Emeneau, Selected and Introduced by Anwar S. Dil*, 1-18. Stanford: Stanford University Press, 1980_2.

Ertelishvili, Parnaoz. *Zmnuri puʒeebis ponemat'uri st'rukt'urisa da ist'oriis sak'itkhebi kartulši. [Phonematic structure and historic aspects of verbal roots in Georgian]*. Tbilisi: Tbilisi State University Press, 1970.

Gersamia, Rusudan. Int'ervok'aluri redup'lik'acia kartvelur enebši [Intervocalic reduplication in Kartvelian languages]. *Issues of Linguistics* 1 (2015): 46-62.

Gil, David. Georgian reduplication and the domain of distributivity. *Linguistics* 26 (1988): 1039–1065.

Holisky, Dee Ann. On the study of expressives in Kartvelian languages. *Proceedings of the First International Symposium in Kartvelian Studies*, edited by Elguja Khintibidze, 52–64. Tbilisi: Tbilisi State University Press, 1988.

Holisky, Dee Ann, & Nana Kakhadze. Manner of Speaking Verbs in Georgian. Part 1: An Exercise in Semantic Description with Comments on Sound Symbolis. In *Studia Caucasiologica I: Proceedings of the Third Caucasian Colloquium, Oslo, July 1986*, edited by Fridrik Thordarson, 184-207. Oslo: Norwegian University Press, 1988.

Hurch, Bernhard (ed.). *Studies on Reduplication*. Berlin & New York: Mouton de Gruyter, 2005.

Inkelas, Sharon, & Cheryl Zoll. *Reduplication: Doubling in Morphology*. Cambridge: Cambridge University Press, 2005.

Jorbenadze, Besarion. *The Kartvelian Languages and Dialects (General Review)*. Tbilisi: Mecniereba, 1991.

Jorbenadze, Besarion, Manana Kobaidze, and Marine Beridze. *Kartuli enis morpemebisa da modaluri element'ebis leksik'oni [Dictionary of Morphemes and Modal Elements of the Georgian Language]*. Tbilisi: Mecniereba, 1988.

Kartozia, Guram, Rusudan Gersamia, Maia Lomia, and Taia Tskhadaia. *Megrulis lingvist'uri analizi [Linguistic Analysis of Megrelian]*. Tbilisi: Meridian, 2010.

Kikvidze, Zaal. Multiple metalanguages of linguistics? Terminology of sound symbolism through time and space. In *Language in Professional Dimension: Communicative and Cultural Aspects*. Part 1, 152-154. Kharkiv: NUCZU, 2014.

Kikvidze, Zaal. Lexicalized and Grammaticalized C_1aC_2 Onomatopes in Georgian: Patterns and Translation Options. In *General and Specialist Translation / Interpretation: Theory, Methods, Practice*, 271-277. Kyiv: Agrar Media Group, 2018.

Kiria, Chabuki, Lali Ezugbaia, Omar Memishishi, and Merab Chukhua. *Lazur-megruli gramatika. 1. morpologia [Grammar of Megrelian-Laz. 1. Morphology]*. Tbilisi: Meridian, 2015.

Kobalava, Izabela. Zogierti t'i'p'is xmabaʒviti sit'q'vebis c'armoebisatvis kartulši [On the formation of onomatopoeic words of some types in Georgian]. *Issues in Contemporary General Linguistics* 5 (1980): 67-75.

Magnus, Margaret. *What's in a Word*? *Studies in Phonosemantics*, PhD thesis. Department of Linguistics, Science and Technology. NTNU Faculty of Arts, Norwegian University, 2001.

Moravcsik, Edith. Reduplicative constructions. In *Universals of human language: Word structure*, Vol. 3, edited by Joseph H. Greenberg, 297–334. Stanford, CA: Stanford University Press, 1978.

Neisser, Friedrich. *Studien zur Georgishen Wortbildung [Studies in Georgian Word-Building]*, Hrsg. von Gerhard Deeters. Wiesbaden: Deutsche Morgenlandische Gesellschaft/Steiner. (Abhandlungen für die Kunde des Morgenlandes; 31, 2), 1953.

Pelliot, Paul. Brèves remarques sur le phonétisme dans l'écriture chinoise**.** *T'oung Pao* 32/2-3 (1936): 162-166.

Pott, Adolf Friedrich. *Doppelung (Reduplikation, Gemination) als eines der wichtigsten Bildungsmittel der Sprache, beleuchtet aus Sprachen aller Welttheile*. Lemgo/Detmold: Meyer, 1862.

Raimy, Eric. *The Phonology and Morphology of Reduplication*. Berlin & New York: Mouton de Gruyter, 2000.

Rayfield, Donald (ed.-in-chief). *A Comprehensive Georgian-English Dictionary*. London: Garnett Press, 2006.

Sagliani, Medea. Int'erpiksuri (resp. "maertxovniani" da "maertmarcvliani") redup'lik'acia svanurši [Interfixal reduplication in Svan]. *Intercultural Communications* 25 (2015): 14-28.

Sanikidze, Latavra. Puʒegaork'ecebuli k'mp'ozit'ebi kartulši [Stem-reduplicated compounds in Georgian]. *Macne* 3 (1968): 144-172.

Sanikidze, Latavra. Gaork'eceba kartulši [Reduplication in Georgian]. *Annual of Ibero-Caucasian Linguistics* 3 (1976): 35-86.

Sanikidze, Latavra. Redup'lik'acia svanursa da zanurši [Reduplication in Svan and Zan]. *Annual of Ibero-Caucasian Linguistics* 4 (1977): 22-80.

Shanidze, Akaki. *Kartuli enis gramat'ik'a I (morpologia) [A Grammar of the Georgian Language I (Morphology)]*. Tbilisi: Tbilisi State University Press, 1955.

Stolz, Thomas, Cornelia Stroh, and Aina Urdze. *Total Reduplication. The Areal Linguistics of a Potential Universal.* Berlin & Boston: Akademie Verlag, 2011.

Tandilava, Ali. *Lazuri leqksik'oni [A Dictionary of Laz]*. Tbilisi: Saari, 2013.

Topuria, Varlam. Kartvelur enata sit'q'vac'armoebidan, VI, k'ompozit'I [From word-formation of Kartvelian languages, VI, Compound]. In *Works* by Varlam Topuria, Vol. 3, 103-131. Tbilisi: Mecniereba, 1979.

Tschanischwili, Nani. Zur semantischen Farbung der Laute im Neugeorgischen [On the Semantic Coloring of Sounds in the Modern Georgian Language]. *Wissenschaftliche Zeitschrift der Friedrich Schiller Universitat Jena, Gesellschafts / Sprachwissenschaftliche Reihe* 37/2 (1988): 167-179.

Voronin, Stanislav S. *Osnovy fonosemantiki [Foundations of Phonosemantics]*. Leningrad: Leningrad University Press, 1982.

Zhgenti, Sergi. *Kartvelur enata šedarebiti ponet'ik'a, I [Comparative Phonetics of Kartvelian Languages, I]*. Tbilisi: Tbilisi State University Press, 1960.

LingBaW *Linguistics Beyond And Within 4 (2018), 108-128*

Laying the foundations: A bibliometric analysis of L2 vocabulary research in 1982-1986

Paul Meara
Swansea University, the United Kingdom

Abstract

This paper uses a co-citation analysis to examine the research on L2 vocabulary acquisition that was published in 1986. This year seems to mark a serious consolidation of L2 vocabulary research, with the main themes of future research appearing. The paper also reports a larger analysis of all the work that appeared in a five-year window from 1982-1986. This analysis clearly shows the beginnings of a recognisable research focus on L2 vocabulary acquisition, though this work is influenced by some surprising sources, who do not figure in more recent work in the field.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric analysis

1. Introduction

This paper is the sixth in a series of studies which attempt to plot the way research in L2 vocabulary acquisition has progressed over the last fifty years. Earlier papers in this series have analysed the research outputs published in 1982, 1983, 1984, 1985 and 2006. (Meara 2012, 2014, 2015, 2016, and 2017). This paper follows on directly from my analyses of the 1983, 1984 and 1985 data, published in earlier issues of *LingBaW* in that it covers the 1986 research output. However, the earlier analyses now provide us with a large amount of data from a five year window, and we can combine the five single-year analyses into a much larger – and more reliable – data set covering the whole of the period 1982-86.

This paper, then, falls into two parts. Part 1 reviews the research published in 1986 in its own terms, while Part 2 analyses the entire research output for the period for 1982-86. Both sections use the co-citation methodology which will by now be familiar to readers of *LingBaW*. The methodology is summarised in Appendix 1 for readers who are not yet familiar with this approach.

2. Part 1. The 1986 data

The main feature to note from our earlier analyses is that 1985 seems to mark some kind of a watershed in L2 vocabulary research. This is the first year in which we find a coherent L2 vocabulary research initiative in the data – a clear contrast with earlier years, where the L2 research seems to rely on a wide range of disparate influences. The 1985 research draws heavily on the work of an active group of European researchers, and is particularly influenced by the work of Scandinavian researchers (notably Håkan Ringbom) and an emerging group of researchers active in the Netherlands. We also noted a continuing rift between research that is informed by psycholinguistics, and research that is more allied with research that comes from an Applied Linguistics tradition, particularly the Edinburgh School with its interest in Error Analysis. Researchers from these two traditions are only rarely cited along-side each other. We also noted the appearance of French and German sources alongside the better known English language research, and the growth of an influential group of sources based in Israel.

The question we need to ask of the 1986 data is whether these trends consolidate: Is the "new beginning" that we identified in 1985 really the start of modern research in L2 vocabulary acquisition, or is it no more than a brief flash in the pan?

2.1. The data sources

A total of 98 research outputs were identified in the VARGA database for 1986 (Meara n.d.) – slightly more than the number of outputs identified in 1985. These outputs include four Masters' theses, one PhD thesis (Quigley), three book length treatments that are principally concerned with teaching materials (Crow, Gairns and Redman, Morgan and Rinvolucri), one monograph that deals with L2 learners' use of words (Linnarud), and a second monograph that deals with the vocabulary of L2_German speakers in primary schools (Neuner and Schade). We also have a small number of conference papers and one unpublished (but influential) report by Black. Conventionally, theses and other book length treatments are excluded from bibliometric analyses on the grounds that the way they cite previous research is very different from the way this work is normally cited in research papers. This practice is followed here. Additionally, a small number of sources turned out to be unobtainable. This left a total of 81 eligible outputs to be used in the analysis, a slight fall from the 1985 figure. These outputs are listed in Table 1. The excluded items are listed in Appendix 2.

Table 1: *The outputs used in the analysis.*

Allwood, J and E Ahlsén
Lexical convergence and language acquisition. In: Ö Dahl (ed.) *Papers from the Ninth Scandinavian Conference of Linguistics Stockholm*: University of Stockholm, Dept of Linguistics. 1986.
Amer, A
Semantic field theory and vocabulary teaching. *English Teaching Forum*, 24,1(1986), 30-33.
Arnedt, CS and JR Gentile
A test of dual coding theory for bilingual memory. *Canadian Journal of Psychology* 40,3(1986), 290-299.
Balcke, D
Wortschatzsystematisierung im Englischunterricht. *Fremdsprachenunterricht* 30,1(1986), 34-35.

Barnett, M
Syntactic and lexical/semantic skill in foreign language reading: importance and interaction. *Modern Language Journal* 70,(1986), 343-349.
Baumgarten, H
Anregungen zum Einsatz der Lautschrift als Hilfe zur schriftlichen Bewaeltigung des Wortschatzes. *Englisch* 21,2(1986), 55-56.
Beheydt, L
Optimalisering van de woordenschatverwerving. [Optimising vocabulary acquisition.] Levende Talen, 416(1986), 630-637.
Bogaards P
Op zoek naar nuttige woorden. Een geschiedenis van de vocabulaireselectie voor het vreemde-talenonderwijs. [In search of useful words.] *Levende Talen* 416(1986), 626-629.
Bogaards, P
Hoe nuttig zijn frequente woorden? [How useful are frequent words?] *Levende Talen* 416(1986), 626-629.
Broeder, P, J Coenen, G Extra, R van Hout and R Zerrouk
Ontwikkelingen in het nederlandstalig lexicon bij anderstlige volwassen: Een mcro- en micro-perspectif. In: **J Creten, G Geerts and K Japaert** (eds.) *Werk-in-uitvoering. Moment-opname van de sociolinguistiek in Belgie en Nederland.* Leuven-Amersfoort: Acco. 1986. 39-60.
Broeder, P, G Extra and R van Hout
Acquiring the linguistic devices for pronominal reference to persons: a cross-linguistic perspective on complex tasks with small words. In: **F Beukema and A Hulk** (Eds). *Linguistics in the Netherlands* Dordrecht: Foris. 1986.
Broeder, P and K Voionmaa
Establishing word class distinctions in the vocabulary of adult language learners; a cross linguistic perspective. In: **O Dahl** (Ed). *Papers from the ninth Scandinavian conference of linguistics*. Stockholm: Dept. of Linguistics. 1986.
Cammarota, M and J Giacobbe
L'acquisition du lexique en français par des adultes hispanophones. [How adult Spanish-speakers acquire vocabulary in French.] *Langages* 84(1986), 65-78.
Chen, HC and C Ho
Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition* 12(1986), 397-401.
Clarke, D
Vocabulary acquisition and the computer database. *UEA Papers in Linguistics*, Special Issue (1986), 21-42.
Cohen, A
Forgetting foreign language vocabulary. In: B Weltens, K de Bot and T van Els (Eds.) *Language Attrition in Progress*. Dordrecht: Foris. 1986.
Cohen, A
Verbal report as a source of information on reading strategies. *ESPecialist* 15,1(1986), 1-12.
Cristoffanini, P, K Kirsner and D Milech
Bilingual lexical representation: the status of Spanish-English cognates. *Quarterly Journal of Experimental Psychology* 38A (1986), 367-393.
Crow, J
Receptive vocabulary acquisition for reading comprehension. *Modern Language Journal* 70,3(1986), 242-250.
Daams-Moussault, A
Woordverwerving en lesmateriaal. *Levende Talen* 416(1986), 644-649.
de Bot, K
Fonologische processen in woordherkenning door twee-talige kinderen. [Phonological processes and word recognition in bilingual children.] *Interdisciplinaire Tijdschrift voor Taal- en Tekstwetenschap* 6(1986), 339-354.
Dorriots, B
How to succeed with only fifty words: an analysis of role-play in the frame of adult language acquisition. *Gothenburg Papers in Theoretical Linguistics*, 52. 1986.

Fichez-Vallez, E
Le vocabulaire a-t-il sa place dans la formation des adultes a l'expression? *Langue et Institutions* 1986. 181-213.
Fox, J
Past, present and future in computer assisted vocabulary learning. *Zielsprache Englisch* 16,3(1986), 25-30.
Gebhard, I
Behaltensfoerderndes Einpraegen und Einueben der Lexik nach Empfang des Fernsehkurses. *Fremdsprachenunterricht* 30,6(1986), 268-270.
Giacobbe, J and M Cammarota
Learners' hypotheses for the acquisition of lexis. *Studies in Second Language Acquisition* 8(1986), 327-342.
Giacobbe, J and M Cammarota
Un modèle de rapport langue source/langue cible dans la construction du lexique. [How the relationship between L1 and L2 affects the development of a lexicon.] In: **A Giacomi and D Véronique** (Eds.) A*cquisition d'une langue étrangère*. PUP. 1986. 192-209.
Giacobbe, and P Ostyn
Apprendre … communiquer en langues étrangères. [Learning to communicate in foreign languages.] *IRAL* 24,4(1986) 329-336.
Graham, C and R Kirk Belnap
The acquisition of lexical boundaries in English by native speakers of English. *IRAL* 24,4(1986), 275-286.
Green, D
Control, activation and resource: a framework for the control of speech in bilinguals. *Brain and Language* 27(1986), 210-223.
Greidanus, T, P de Beyl and J Schouwerwou
Het opmaken van de betekenis van onbekende woorden uit de context. [Guessing the meaning of unknown words in context.] *Levende Talen* 416(1986), 638-644.
Hakansson, G
Quantitative aspects of teacher talk. In: **G Kasper** (ed.) *Learning, teaching and comunication in the foreign language classroom*. Aarhus: Aarhus University Press. 1986. 83-98.
Havola, S and S Takala
Review of some recent soviet research on vocabulary learning and teaching. *Papers in Linguistics* 19(1986), 495-510.
Heid, U
Wortschatzlernen mit dem Computer. Was taugen die kommerzialisierten Vokabeltrainingsprogramme? [Learning words with the computer.] *Zeitschrift für Sprache un Literatur in Wissenschaft und Unterricht* 58(1986), 68-84.
Höhmann, HO
Lexikalische Lernarbeit mit dem Lehrbuch. *Der fremdsprachliche Unterricht* 20,79(1986), 171-180.
Holmes, JL
Snarks, quarks and cognates: an elusive fundamental particle in reading comprehension. *ESPecialist* 15(1986), 13-40.
Huckin, T
The use of discourse patterning in foreign language reading and vocabulary acquisition. *Documentos de Estudos em Linguística Teórica e Aplicada* 2,1(1986), 57-75. Sao Paulo.
Huckin, T and Zhendong Jin
Inferring word-meaning from context: a study in second language acquisition. In: ESCOL '86. *Proceedings of the 3rd Eastern States Conference on Linguistics.* 1986. 271-280.
Ijaz, H
Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning* 36,4(1986), 401-451.
Johns, T
Micro-concord: a language learners' research tool. *System* 14,2(1986) 151-162.

Kellerman, E
An eye for an eye: Crosslinguistic constraints on the development of the L2 lexicon. In: **E Kellerman and M Sharwood Smith** (eds.) *Crosslinguistic Influence in Second Language Acquisition*. Oxford: Pergamon Press. 1986. 35-48.
Kelly, P
Solving the vocabulary retention problem. *ITL Review of Applied Linguistics* 74(1986), 1-16.
Kirsner, K
Lexical function: is a bilingual account necessary? In: **J Vaid** (Ed.) *Language processing in bilinguals*. Hillsdale, NJ.: Lawrence Erlbaum Associates. 1986.
Kohlert, I
Arbeit mit individuell-differenzierter Lexik im Franzoesischunterricht. *Fremdsprachenunterricht*, 30,6(1986), 283-284.
Laufer, B
A case for vocabulary in EAP reading comprehension. In: **A-M Cornu, J Vanparijs, N N Delahaye and L Baten** (Eds.) *Beads or Bracelet? How do we approach LSP?* Oxford: Oxford University Press. 1986.
Laufer, B
Possible changes in attitude towards vocabulary acquisition research. *IRAL* 24,1(1986), 69-75.
Lessard, G
The presentation of vocabulary in computer-aided instruction. *Canadian Modern Language Review* 43(1986), 94-107.
Lübke, D
Vokabellernen mit Computer? *Praxis des deustchsprachlichen Unterrichts* 4(1986), 387-394.
Mack, M
A study of semantic and syntactic processing in monolinguals and fluent early bilinguals. *Journal of Psycholinguistic Research* 16,6(1986), 463-488.
Mahnert, D
Worte, Worte – nichts als Worte! Vokabeln einführen, üben und lernen in der Sekundarstufe I. *Der fremdsprachliche Unterricht* 20, 1986, 79, 191-202.
Meara, PM
The DIGAME project. In: **VJ Cook** (Ed.) *Experimental approaches to second language learning*. Oxford: Pergamon Press. 1986. 101-110.
Meara, PM
Psycholinguistic relativity – is Spanish different? In: **VJ Cook** (Ed.) *Experimental approaches to second language learning*. Oxford: Pergamon Press. 1986.
Meara, PM and S Ingle
The formal representation of words in an L2 speaker's lexicon. *Second Language Research* 2(1986), 160-171.
Murray, DJ
Characteristics of words determining how early they will be translated into a second language. *Applied Psycholinguistics* 7,4(1986), 353-372.
Nakamura, LK
Students' identification of English/Portuguese cognates. *ESPecialist* 13(1986), 11-18.
Nakamura, LK
A practical procedure to develop student's awareness of English imports into the language. *ESPecialist* 13(1986), 19-24.
Nattinger, J
Lexical phrases, functions and vocabulary acquisition. *The ORTESOL Journal* 7(1986) 1-14.
Noyau, C and M-T Vasseur
L'acquisition des moyens de la référence temporelle en français langue étrangère chez des adults hispanophones. [Acquisition of terms for time in French by native speakers of Spanish.] *Langages* 84(1986), 105-117.

Ostyn, P, M Vandecasteele, G Deville and P Kelly
Towards an optimal programme of FL vocabulary acquisition. In: **A-M Cornu, J Vanparijs, M Delahaye & L Baten** (Eds). *Beads or bracelet? How do we approach LSP?* Leuven: Instituut voor Levende Talen. 1986.
Palmberg, R
Vocabulary teaching in the foreign language classroom. *English Teaching Forum* 24, 3(1986), 15-24.
Pattison, P
'Je leert je rot aan woorden.' [School-children's attitude towards vocabulary learning.] *Levende Talen* 416(1986), 649-653.
Pons-Ridler, S
Les champs semantiques de l'action a travers le vocabulaire fondamental [Semantic Fields and Verbs in the Dictionary of Basic Vocabulary]. *French Review* 59,4(1986), 515-531.
Pye, C
One lexicon or two? an alternative interpretation of early bilingual speech. *Journal of Child Language* 13(1986), 591-193.
Ramirez, SZ
The effects of suggestopedia in teaching English vocabulary to Spanish-dominant Chicano third graders. *The Elementary School Journal* 86,3(1986), 324-333.
Reiner, E
Dubletten und Faux Amis: Berührungspunkte zwischen Sprachunterricht und Sprachwissenschaft. *Moderne Sprachen* 30,3-4(1986), 30-48.
Rudzka-Ostyn, B
Vocabulary teaching: a cognitive approach. In: **A-M Cornu, J Vanparijs, N Delahaye and L Baten** (Eds.) *Beads or Bracelet? How do we approach LSP?* Oxford: Oxford University Press. 1986.
Schouten-van Parreren, C
Nieuwe perspectieven op de didactiek van de woordenschatverwerving. [New perspectives on the teaching of vocabulary]. *Levende Talen* 416(1986), 618-625.
Schouten-van Parreren, C
Verslag van de werkbijeenkomst woordenschatuitbreiding. *Verslag* (1986) 143-148.
Schumans, J and W Hermans
Woordkennis Duits in HAVO3. [Knowledge of German vocabulary in Dutch Secondary Schools]. *Toegepaste Taalwetenschap in Artikelen* 26(1986), 95-102.
Segalowitz, N
Skilled reading in the second language. In: **J Vaid** (ed.) *Language processing in bilinguals: Neuropsychological perspectives.* Hillsdale, NJ.: Lawrence Erlbaum Associates. 1986. 3-19.
Slagter, PJ
Woorden, woorden, woorden. [Words, words, words.] *Levende Talen* 416(1986), 616-617.
Smit-Kreuzen, M
Woordenschatverwerving – kan de computer helpen? [Vocabulary acquisition: can the computer help?] *Levende Talen* 416(1986), 648-651.
Speight, S
Choosing the right words. *Der fremdsprachliche Unterricht* 20, 1986, 79, 182-190.
Stip, P and J Hulstijn
Hoe geef je het goede voorbeeld? Woordenschatuitbreiding met behulp van voorbeeld zinnen. [Explaining word meanings with target language example sentences.] *Toegepaste Taalwetenschap in Artikelen* 25(1986), 118-128.
van Kooten, WA
Vocabulariumverwerving en vocabulariumbeheersing op universitair niveau: Focus on the learner. *Toegepaste Taalwetenschap in Artikelen* 25,1(1986), 110-117.
Véronique, D and R Porquier
Acquisition des moyens de la référence spatiale en français par des adultes arabophones et hispanophones. [How native Arabic and Spanish speakers acquire terms for spatial reference in French.] *Langages* 84(1986) 79-103.

Vihman, M
More on language differentiation. *Journal of Child Language* 13(1986), 595-597.
Weis, D
Untersuchungen zur langfristigen Verfugbarkeit von Wortschatz im Leistungsfach Englisch. *Neusprachliche Mitteilungen* 3(1986), 174-180.
Wolf-Franz, J
Was ich mit meinem Koerper alles kann – Wortschatzerweiterung und Wortschatzsicherung. *Grundschulmagazin* 1,12(1986), 11-12.
Zimmerman, R
Semantics and lexical error analysis. *Englisch, Amerikanische Studien* 2(1986), 294-505.
Zimmerman, R
Classification and distribution of lexical errors in the written work of German learners of English. *Papers and Studies in Contrastive Linguistics* 21(1986), 21-41.

The most notable feature of the included items is that many of them are to be found in two special issues of the Dutch language journals *Levende Talen* and *Toegepaste Taalwetenschap in Artikelen* – a clear confirmation of the importance of the Dutch research that we noted in the 1985 analysis.

The usual superficial analysis of the number of contributions made by each author in the data set identifies 92 unique contributors (again, slightly down from the 1985 figure).

As in previous years, the number of authors who contribute to a single paper is large (83% of the total). However, the number of authors contributing to more than one output in the 1986 data set has increased to 16: **Giacobbe** contributed to 4 papers; **Broeder, Cammarota** and **Meara** each contributed to 3 papers; **Bogaards, Cohen, Extra, Huckin, Kelly, Kirsner, Laufer, Nakamura, Ostyn, Schouten-van Parreren, van Hout** and **Zimmerman** all contributed to two papers (see Table 2). The biggest contributor to the 1986 data set was Giacobbe who contributed to 4 outputs in 1986 – the highest figure that we have found so far in this series of reports. As in previous years, the data suggests that there are fewer authors of multiple outputs than we would expect in a mature discipline. Work by Lotka (1926), conventionally known as Lotka's law suggests that we can expect the number of authors contributing N papers to a field will be approximately $1/N^2$ times the number of authors making a single contribution to the field. Table 2 shows that the number of authors contributing 2, 3 and 4 papers in 1986 falls well short of the figures predicted by Lotka's Law: the field as a whole continues to over-rely on one-off contributions.

Table 2: *Authors contributing to N publications in 1986*

No. of contributions	4	3	2	1
No. of Authors making N contributions in 1986	1	3	12	76
Expectation from Lotka's Law when N_1=76	2	8	19	

Of the 16 authors who contributed more than a single paper to the 1986 dataset several had also made multiple contributions in 1985 (**Broeder, Extra, van Hout, Laufer** and **Meara**). **Cohen** and **Schouten-van Parreren** were both identified as significant influences in the 1984 and 1985 data, though they did not make multiple contributions in those years. Two authors had published relevant papers in earlier years, but this work was not cited often enough for them to

reach the inclusion threshold for this study (Bogaards 1980 and Kirsner et al. 1980, 1982 and Smith and Kirsner 1984). Six authors are genuinely new members of the list, who have not appeared in our earlier analyses: **Giacobbe, Cammarota, Huckin, Kelly, Kirsner, Nakamura** and **Zimmerman.**

2.2. The analysis

The citation data from each of the 81 eligible papers was extracted in the usual way (see Appendix 1). This analysis identified 1278 unique sources. The number of times each of these sources is cited in the data set is summarised in Table 3.

Table 3: *The number of times sources are cited in the 1986 data set*

frequency	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	1		1	2		1	1	4	17	26	72	179	974

The ten most cited sources in this data set are: **Krashen(13), Meara (11), Levenston** and **Schouten-van Parreren (10)**, **JC Richards (8)** and **Lambert, Corder, Gougenheim, Hatch** and **Rosch** with four citations each. Compared with the 1985 data, the raw number of sources has increased slightly, and so has the number of sources cited only once in the data set. The main change between 1985 and 1986 is the number of sources who are cited a lot. Krashen, for example, is cited in 16% of the papers in the data set.

This distribution is actually quite difficult to work with. In 1985, 88 sources were cited at least three times in the dataset. In 1986, the equivalent figure is 125 sources, an increase of 42%. 125 sources is quite a lot larger than the conventional figure of 100 often used in bibliometric maps. We might consider using a higher inclusion threshold for the 1986 data, but only 53 sources were cited at least four times, and this figure is very much lower than we would like for a co-citation analysis. In the analysis that follows, then, we will work with the 125 sources that are cited at least three times in the 1986 data. Included authors are therefore cited in about 4% of the total outputs in 1986, and the inclusion criteria for 1986 are identical to the inclusion criteria for 1985.

These data were analysed using the Gephi software (Bastian, Heymann and Jacomy 2009). The 125 sources are linked by 2031 edges, and Gephi identifies four main clusters in the data. The analysis is shown in Figure 1. For the sake of simplicity the weakest links have been eliminated from this graph. The nodes are sized according to their **betweenness centrality**. (Nodes have a high score on this measure if they are likely to be included in the shortest path between two randomly selected nodes. The measure tends to highlight authors who act as bridges between clusters.)

The paragraphs that follow identify the main features of these clusters. We will follow this discussion with a consideration of the sources that survive from 1985-1986, and the new sources that appear in the 1986 map.

Cluster I, at the Western edge of the graph is the by now familiar psycholinguistics group of influences. As usual, the members of this cluster are very frequently cited alongside each other, but they have very limited co-citations with members of the other clusters. Most of these

between cluster co-citations link to Kucera and Francis, the standard word frequency count that both psychologists and applied linguists were using at this time.

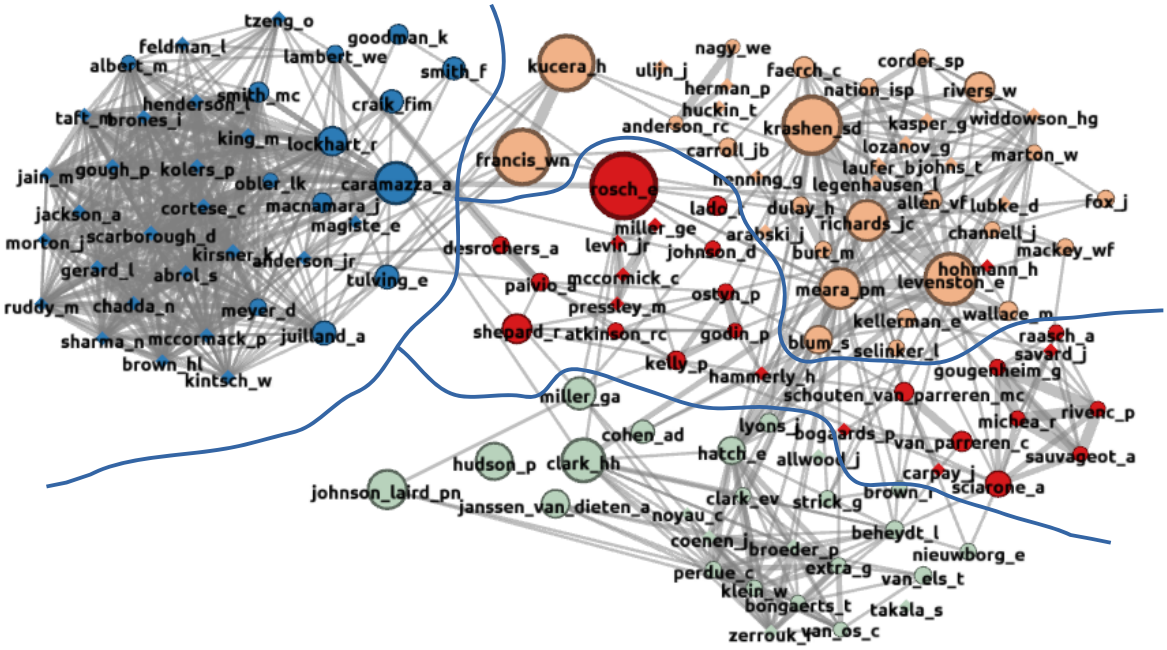


Figure 1: *Co-citation analysis of the 1986 data. Each source is cited at least three times in the data set.*

Cluster II, at the bottom of the map is slightly more difficult to characterise. This cluster too contains some influences who are psycholinguists (notably Johnson-Laird and George Miller), and it also contains a group of influences who are mainly concerned with L1 acquisition (Eve Clark, Herbert Clark and Roger Brown). The main feature of this cluster is the relatively dense part of the map which includes Broeder, Extra, Bongaerts, van Els, Zerrouk, Perdue and Klein. These influences are all members of a large international project team funded by the European Science Foundation, and hosted by the Max Planck Institute in Nijmegen. This project was a very large scale comparative study of the way migrants acquire typologically different languages in naturalistic settings. The project used an interesting methodology which relied heavily on in depth observations of a small number of Subjects across a large number of language pairs – five target languages and six source languages in total. Klein was the overall director of the project. Broeder, Extra and van Hout were part of the Dutch team which looked at acquisition of Dutch by L1 Arabic speakers and by L1 Turkish speakers. Giacobbe and Camarotta, who were identified earlier as authors of multiple papers in the 1986 data set, were part of the French team working on this project, which looked at acquisition of French by L1 Spanish speakers and L1 Arabic speakers. Both teams had a special interest in the lexical development demonstrated by their Subjects, but unlike the reports of the Dutch team, Giacobbe and Camarotta's reports did not appear early enough to influence the other work in the 1986 data set.

Cluster III, at the North East section of the map is easily recognisable as a cluster that deals with mainstream L2 vocabulary research. Interestingly, this cluster is much more tightly interconnected in this map than it was in our earlier maps, and this suggests that there is a growing consensus among L2 vocabulary researchers over who the main influences are. It is

also worth noting that most of the sources co-cited in this cluster are themselves L2 vocabulary researchers, and a very high proportion of them appeared as significant influences in our 1985 map. Perhaps these are the early signs of a self-reflexive orthodoxy emerging in the field?

Cluster IV, in the centre of the map, is the most difficult cluster to characterise. Two main sub-groups stand out among these influences. At the left of this cluster is a group of researchers whose main interests focus on imagery and the role of imagery in memory. Particularly striking here is a group of psychologists who had earlier published a lot of research into the use of mnemonics in vocabulary acquisition (Levin, McCormick, Pressley and Atkinson, whose main work appeared in the late 1970s). The other end of this cluster comprises the *Français fondamental* group (Gougenheim, Michea, Rivenc, Sauvageot and Savard), and three Dutch researchers (van Parreren, Schouten-van Parreren and Sciarone) who are mainly interested in the way learners can infer the meanings of words from context. The single idea that seems to link these separate groups appears to be the specific conditions under which learning words can take place. In this respect, cluster IV looks to be concerned with small-scale vocabulary learning, whereas cluster III might be more concerned with large scale vocabulary learning, with a particular interest in vocabulary use.

Overall, this map differs from the 1985 map in that its structure is much less influenced by geography, and much more influenced by topic. This feels like a significant change.

Several of the sub-clusters that we identified in the 1985 map no longer make an appearance in 1986, confirming our view that the clusters are very volatile at this period.The dictionaries and semantics group no longer stands alone as an identifiable cluster; what remains of this group seems to have been subsumed by Cluster IV. The 1985 dyslexia sub-cluster seems to have disappeared completely. So too has the small cluster comprising Davoust and Bouscaren, which we identified as a specifically French set of influences. The large cluster of influences focussed on word frequency issues seems to have disintegrated in 1986. Most of the members of this cluster still appear in the 1986 map, but here they seem to be distributed among the other clusters, rather than forming a cluster of their own. About half of the members of this cluster have been absorbed by cluster III in the 1986 map, while the French language sources have mostly been absorbed by Cluster IV.

As usual, we can identify a small group of **survivors –** sources who were influential in both the 1985 and the 1986 data – summarised in Figure 2. Because the 1986 network is much larger than the 1985 network, we might expect that the 1986 survivors list would also be larger. Surprisingly, this is not the case. We had 32 survivors in 1985, and 36 in 1986. These sources are shown in Figure 2. In fact, most of the 1986 survivors had appeared as survivors in the 1985 map, yet another indication that the field is beginning to crystallize. A third of the 36 survivors are genuinely new additions to the survivors list: **Bongaerts, Carroll, Hammerly, Hatch, Ostyn, Sciarone,** and the *Français Fondamental* group: **Gougenheim, Michéa, Rivenc, Savard** and **Sauvageot.** It is also worth noting that some of the 1985 survivors, including some sources who would have been considered very influential, no longer appear in the survivors list: **Oller, Ringbom, Clarke, Bialystok, Arnaud, West, Fillmore, Krauthamer** and **Paradis** all seem to have drifted off the radar in 1986. Given that the inclusion criteria for both maps are identical, it seems safe to conclude that some serious consolidation is taking place in 1986, but that some long-standing influences are becoming less significant.

Gephi identifies three clusters in the survivor group. The familiar psycholinguistic survivors appear in the west of the map, but much reduced in number from their dominating role in earlier maps. Apart from Kucera and Francis – the two most important influences in this cluster – only Caramazza is co-cited with influences from the other groups. I think that Caramazza is a ignificant influence in this map because his 1979 paper deals with English and Spanish bilinguals, an interest shared with Meara in the central Cluster.

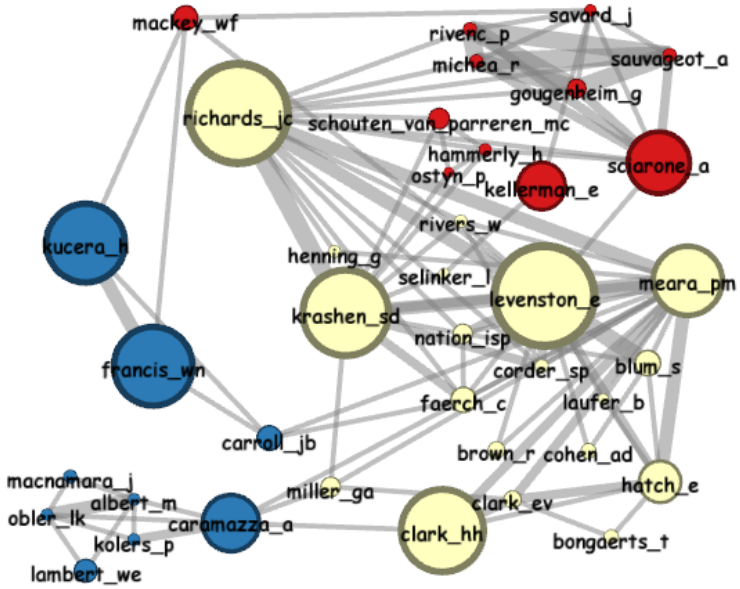


Figure 2: *The 1986 survivors: influences who appear in both the 1985 and the 1986 map. The weakest links have been removed for the sake of simplicity. Nodes are sized according to their betweenness centrality.*

The cluster at the Northeast corner of the survivor map consists of two subgroups, the strongly inter-connected *Français Fondamental* group, and some important influences based in the Netherlands and Belgium. A surprising feature of this cluster is the importance of **Sciarone,** strongly co-cited alongside the *Français Fondamental* group. Sciarone's work is mostly in Dutch, so not well known in the English speaking research world, though his 1979 book is very frequently cited by Dutch researchers.

The centre of the map is occupied mainly by English language researchers, and contains over half of the most significant influences in this map. This central cluster is dominated by Levenston (who together with Laufer, Blum and Cohen make up the Israeli cohort here.) Richards and Krashen make up the other two wings of this influential triumvirate. Herbert Clark, Eve Clark and Roger Brown speak to the continuing influence of L1 acquisition research on L2 vocabulary thinking. The surprising survivors here are Corder and Selinker, not vocabulary researchers themselves, but clearly influencing the thinking of many of the papers in the 1986 data set.

Figure 3 shows a map of the new influences who appear in the 1986 data. As usual, the weakest links are omitted in the interests of simplicity and the the nodes are sized according to their betweenness centrality.

This map needs to be treated with rather more caution than the maps in our earlier papers that identified sources surviving from one year to the next. type. At first glance, this map seems

to imply that we have an unusually large number of new influences in 1986. In part, however, this increase is due simply to the fact that the 1986 map contains many more nodes than the 1985 map, and this inevitably leaves room for a lot of new entrants. Nevertheless, the criteria for inclusion in the 1985 map and the 1986 map are identical, so this large increase in new entrants does seem to be a genuine phenomenon, and not just an artefact. And there are some striking features in the 1986 data which require comment.

Gephi identifies four clusters in this data. These clusters broadly reflect the cluster structure of the whole 1986 data set – this is not really surprising given that new entries make up a very large proportion of the bigger map. The new clusters are rather more interconnected than we have come to expect from our earlier maps where the new entrants formed sizeable but loosely connected clusters. Here, each cluster is fairly self-contained, but there are few strong ties running across the cluster boundaries.

The most striking feature of this map is the very dense cluster that appears at the Western edge of the map, a cluster that is largely made up of psychologists and psycholinguists. A feature of this sort has appeared in most of the maps in this series of papers. What is surprising here, however, is the sheer number of new sources appearing in this cluster. It implies that the psycholinguistics cluster which appears in our earlier maps is not nearly as stable as it appears at first sight: the 1985 map contained a very large psycholinguistics cluster, but by 1986 most of its members have been replaced by new influences. The new theme in 1986 seems specifically to be word recognition in an L2, a theme which has not been strongly represented in the earlier maps. Several of the papers included in the 1986 data set come from a single volume of papers edited by Vaid, and this may have influenced the emergence of a new and unusually coherent theme in the 1986 map.

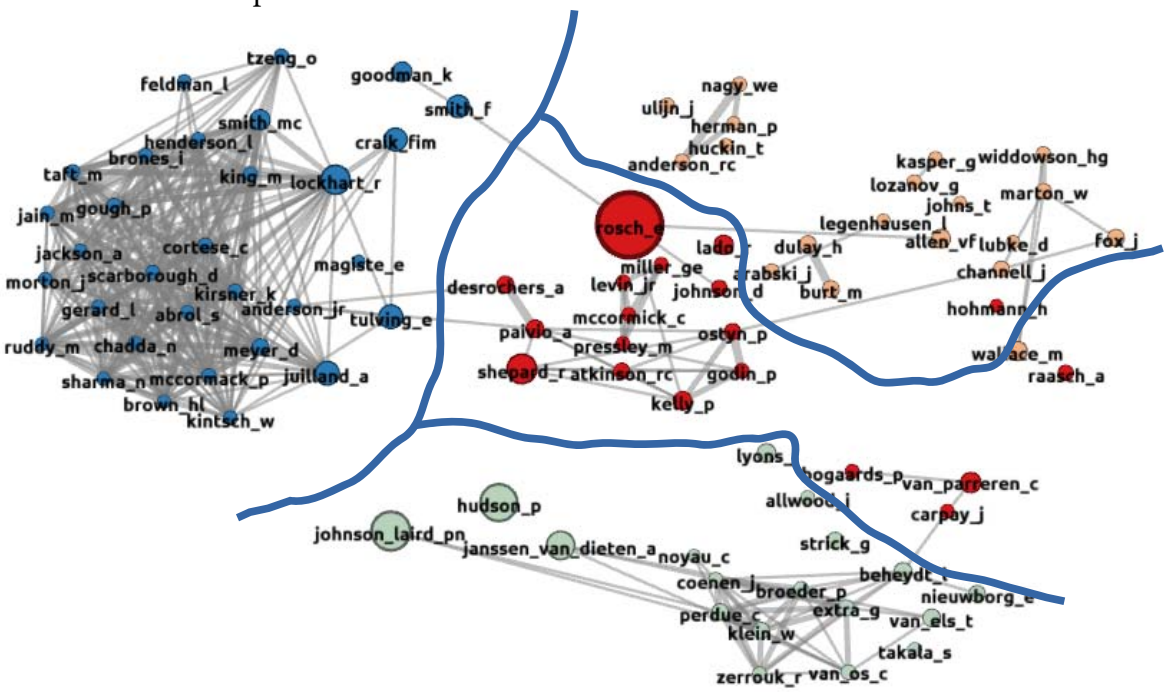


Figure 3: *The new influences in the 1986 map.*

The cluster at the Southern edge of the map is largely made up of the members of the ESF project, and a couple of influences in the area of semantics. Lyons and Beheydt were not

significant sources in the 1985 map, but did play a role in earlier maps, suggesting that they are returners, rather than genuinely new sources. This cluster, despite its strong internal structure, is largely isolated from the rest of the map: it has no co-citation links with two of the other clusters, and only a weak geographical link with the remaining cluster.

The cluster at the Northeast sector of the map contains two sub-clusters. The Nagy-Huckin-Anderson-Herman-Ulijn sub-cluster represents a set of influences who are mainly concerned with L2 reading skills. (We might have expected them to be closely linked with Goodman and Smith in the psycholinguistics cluster – both eminent figures in L1 reading studies.) The remaining members of this cluster seem to represent a set of eclectic approaches towards vocabulary acquisition. Lozanov (of Suggestopedia fame) is a notable figure in this group, Johns and Fox represent the beginnings of a computer-based approach to vocabulary acquisition. Dulay and Burt reflect the increasing dominance of Krashen in the US research. Allen and Wallace both authored text books on vocabulary teaching.

Cluster IV, the group at the centre of this map, is also difficult to characterise succinctly. Rosch has appeared as a significant influence in our earlier maps, but in spite of her central role in the overall graph, she seems to be somewhat disconnected from the other influences in this cluster. Two sub-groups can be identified. The van-Parreren-Carpay-Bogaards subgroup is a nucleus of Dutch vocabulary researchers. The remaining members of this cluster are very specifically interested in imagery and mnemonics as they apply to L2 vocabulary acquisition.

Apart from the striking turn-over in Cluster 1, the new entrants seem to reinforce the general patterns which have appeared in our earlier maps. By 1986, we still do not have anything like a consensus about the main lines of vocabulary research, and the clusters of influences which inform vocabulary research are very fluid and unsettled. Nonetheless, a small core of influences, whose names will be familiar to modern researchers, is beginning to appear consistently in these maps. None of these figures has yet emerged as a dominant influence on the way people are thinking about vocabulary at the time, however. That role is reserved for figures like Krashen and Rosch, neither of whom would be regarded as major L2 vocabulary researchers today. Kucera and Francis are the other major influences in 1986: they play a role which would probably surprise them as much as it is likely to surprise modern readers. Their influence is more methodological than conceptual, with their word frequency list playing an important reference role in discussions of word frequency.

In short, the map which emerges from the 1986 literature continues to look very different from the kind of maps which we find in the more recent research (Meara 2012). The real change here is the fact that more systematic research is being carried out into L2 vocabulary work, and that the multiple papers generated by a couple of coherent research groups are beginning to form a body of work that L2 vocabulary research can start to build on. Perhaps 1986 is best seen not so much as a great leap forwards, but as a substantial amount of preparatory work on the foundations of future research.

3. Part 2. A wider perspective: 1982-1986

Several reviewers of my previous papers have commented that it is rather unusual to carry out bibliometric analyses on a single year's work, and that it might make more sense to work with

a wider window, say, five or ten years' worth of research outputs. This is a very valid comment, which needs to be taken seriously. As we have seen, the annual maps that we looked at so far show quite radical changes from one year to the next, suggesting that the field was peculiarly volatile at this time. However, it is possible that the volatility is an artefact caused by the relatively slow rate of publication at the time. And maybe a more stable picture, one less influenced by short-term publication patterns, would emerge if we stood back from the data and analysed a larger set of papers produced over a longer time period.

Obviously, now that we have data sets from 1982-1986, we are in a position to merge the annual data sets into a much larger comprehensive data set covering the entire period. A bibliometric analysis that does this is presented in this section.

3.1. The data sources

The combined data set for 1982-1986 consists of 317 papers. For reasons of space. I have not listed these papers here, but interested readers can find the included papers in the VARGA database: (www.lognostics.co.uk/varga). Items included in this analysis are listed with two hash marks (##), and the list of included items can be retrieved by using the VARGA query page to limit the search period (1982-1986) and entering ## as the search term. Note that this figure of 317 items is somewhat larger than the combined total of the individual 1982-1986 analyses reported earlier. This is because a few unobtainable items came to light after publication of the earlier analyses. These additional sources are fairly obscure sources which do not materially change the earlier analyses.

The data set identifies a total of 309 unique authors, distributed as shown in Table 4.

Table 4: *Authors contributing N papers to the combined data set, and the number of contributors we would expect taking Lotka's Law into account.*

# contributions	11	10	9	8	7	6	5	4	3	2	1
authors	1					4	1	7	12	42	242
Lotka's Model	2	2	4	4	5	7	10	15	27	60	

The most prolific authors in this period are Meara (11 items), Laufer, Mägiste, Nation and Ringbom (6 items each), Broeder (5 items), Binon, AD Cohen, Cornu, Extra, Kirsner, Palmberg and van Hout (4 items each).

In my earlier analyses, I pointed out that the field as a whole is characterised by a relatively large number of authors who contribute to only a single paper in the data set. This feature is even more apparent in the combined data set than it is in the data for individual years. The bottom line of Table 4 shows how many people we would expect to be publishing N papers in this period, given that we have 242 authors contributing to just one paper (Lotka 1926). Lotka's model suggests that the number of people contributing to N papers is about half of what we would expect, and considerably worse than this in the 7-10 range.

3.2. The analysis

Table 5 shows our initial analysis of the citation patterns in this data. The 15 most cited sources are Lambert (40), Meara (34), Richards(30), Krashen (29), Albert, Corder and Levenston (25 each), Obler(24), EV Clark, AD Cohen and Kolers(23 each), Macnamara(22) and H Clark, Kucera and Francis (21 each). A raw frequency count of the citation data suggested that the threshold for inclusion in the co-citation analysis that follows should be set at 9 citations. This threshold gives a total of 92 sources – very close to the traditional figure of 100 that is conventionally used for co-citation analyses. Figure 4 shows a more detailed analysis of the citation patterns in the data set; the analysis is based on the 370 co-citation links between nodes that occur at least 4 times in the data set. At this level of delicacy, Gephi's analysis identifies 5 clusters and one unattached singleton. These clusters are identified below in order of their size.

Table 5: *The number of authors cited in N papers in the combined 1982-86 data set.*

N cites	**45**	**44**	**43**	**42**	**41**	**40**	**39**	**38**	**37**	**36**	**35**	**34**	**33**	**32**	**31**
sources						1						1			
N cites	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**
sources	1	1				3	1	3	1	3	1	1		3	3
N cites	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
sources	5	5	9	6	4	21	18	31	30	56	86	122	234	530	2201

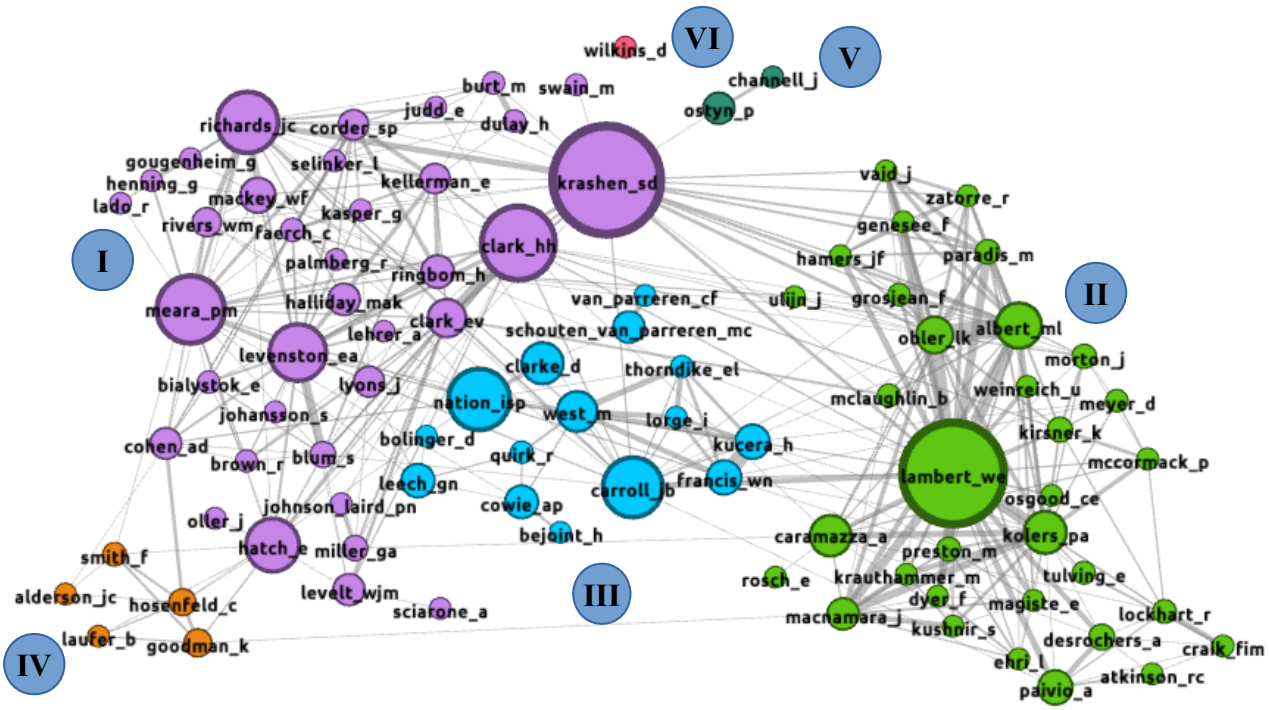


Figure 4: *Patterns of co-citation among the 92 most frequently cited sources in the 1982-86 data set. Threshold for inclusion is 9 citations in the data set, with a minimum co-citation strength of 4. Nodes are sized according to their betweenness centrality.*

Cluster I, the principal cluster with 35 members, can clearly be identified as an L2 vocabulary cluster. The most important influence here is Krashen, but six of the top ten sources (in terms of betweenness centrality) belong to this cluster – Krashen, Clark, Meara, Richards, Levenston

and Hatch. Several subclusters can be identified in this group: *français fondamental* (Gougenheim, Richards, Mackey); semantics and meaning (Lyons, Lehrer, Halliday and Hatch); L1 acquisition (Clark and Clark and Brown); transfer and lexical errors (Ringbom, Corder, Selinker, Kellerman, Palmberg); a psycholinguistics influence (Miller, Johnson-Laird, Levelt); and the Krashen, Dulay and Burt subcluster, mainly identified with Krashen's Monitor Model, and the difference between acquisition and learning. Krashen's surprising appearance here as the Most Significant Influence will be discussed in more detail below.

Cluster I, the principal cluster with 35 members, can clearly be identified as an L2 vocabulary cluster. It consists of 35 influences. The most important influence here is Krashen, but six of the top ten sources (in terms of betweenness centrality) belong to this cluster – Krashen, Clark, Meara, Richards, Levenston and Hatch. Several subclusters can be identified in this group: *français fondamental* (Gougenheim, Richards, Mackey); semantics and meaning (Lyons, Lehrer, Halliday and Hatch); L1 acquisition (Clark and Clark and Brown); transfer and lexical errors (Ringbom, Corder, Selinker, Kellerman, Palmberg); a psycholinguistics influence (Miller, Johnson-Laird, Levelt); and the Krashen, Dulay and Burt subcluster, mainly identified with Krashen's Monitor Model, and the difference between acquisition and learning. Krashen's surprising appearance here as the Most Significant Influence will be discussed in more detail below.

Cluster II is the familiar set of psychological influences, which we have noted in our earlier analyses. The influences in this cluster are mainly psychologists interested in experimental studies of bilingual behaviour. Two of our top ten influences fall into this cluster: Lambert and Albert. Again, we can distinguish a several sub-clusters here, principally a number of people interested in neurolinguistics (Albert, Obler, Paradis), a group interested in the linguistic performance of bilinguals (Lambert, Macnamara, Kolers), and a group of influences that are specifically interested in word recognition effects in bilingualism. The distinguishing features of this cluster of influences are a focus on experimental research methods (as opposed to the less rigorous, descriptive methods employed by the researchers in cluster I), and a Montreal connection – several of the influences in this cluster are based in that city. The cluster has relatively few co-citation links with Cluster 1, and almost all of these links involve a co-citation with Krashen. Co-citation links within the cluster are very strong.

Cluster III, with 15 members, is predominantly made up of the main frequency counts available at this time. Thorndike and Lorge, Kucera and Francis, Carroll and West all authored important word counts. Quirk and Leech are important in corpus linguistics. Bejoint and Cowie work on L2 learners and dictionaries. The most interesting influence in this cluster is Nation, forming with Clarke, Schouten-van Parreren and van Parreren, a subcluster that deals with guessing and inferencing behaviour in L2 learners. This cluster has a lot of weak co-citation links with Cluster I, but only one stronger co-citiation link with Cluster II (Carroll~Lambert).

Cluster IV, a small cluster with only five members, is clearly an L2 reading cluster. The main figure of interest here is Laufer, who will become a very significant influence in later maps. Goodman and Smith are both major figures in L1 reading research. This cluster is weakly connected to cluster I, and has one direct co-citation link with Cluster II (Goodman~Macnamara), but no direct co-citation links with Cluster III.

Cluster V contains only two members, Ostyn and Channell, who were working at this time on a semantic approach to vocabulary acquisition. We might have expected to find this cluster closely aligned with the semantics group in cluster I (Lyons, Lehrer, Halliday), but this appears not to be the case.

Finally, Wilkins appears as an unconnected singleton in this map (**Cluster VI**). He is cited nine times in the data set, but none of the co-citation links are strong enough to reach the threshold for inclusion for this map. Wilkins is mostly cited in connection with a comment in his 1972 book:

> 'There is not much value in being able to produce grammatical sentences if one has not got the vocabulary that is needed to convey what one wishes to say ... While without grammar very little can be conveyed, without vocabulary nothing can be conveyed'
>
> (Wilkins 1972:110-111).

4. Discussion

A number of features in the 1982-1986 map deserve some comment. Firstly, our analysis of the five year data set broadly supports our earlier analyses based on the outputs in a series of single years. However, the patterns in the data are much more clearly discernible here, and it does look as though a five year window gives a more reliable and more stable picture, which is less susceptible to the accidental fluctuations that we find in the smaller, single year data sets. We will work with data sets from a rolling five year window in future analyses.

Secondly, the five year data shows a much more coherent picture of the applied linguistics research on L2 vocabulary than we found in the single year analyses. In the earlier analyses, the applied linguistics influences were relatively insignificant compared with the psychological influences in the research. Here, they emerge as considerably more numerous than the psychological influences, more clearly delineated into identifiable research themes. Whereas the L2 vocabulary researchers were relatively insignificant in the single year analyses, here they appear as a large group (Cluster I), and if we add in Cluster 3, Cluster 4, Cluster 5 and Cluster 6 to Cluster 1, then the linguistic influences significantly outnumber the psychological ones. Cluster 1, the main L2 vocabulary acquisition cluster, has not yet crystallised around a single theme, and references sources from a range of research traditions. This is probably what we would expect to find in a relatively young field of research, and it suggests that we might expect a hardening of this group in the years to come, with some of the more tangential influences forming a focus for new clusters. For instance, it would not be surprising to find in future a young bilinguals cluster, heavily influenced by L1 vocabulary research, that focussed on vocabulary acquisition in young bilinguals.

Thirdly, the five year window strongly reinforces our view that the psycholinguistic influences and the applied linguistics influences form two very separate clusters that do not really interact with each other. Cluster II exhibits the strongest co-citation links of all the clusters in this map, but few of these links extend into other clusters. Again, this feature was one that we noted in connection with our single year maps, but the longer term view provided in Figure 4 clearly shows that the majority of the influences in Cluster II have no co-citation links

at all with the other clusters. The gulf between experimental and observational approaches to L2 vocabulary appears to be deeply engrained in this map.

Fourthly, a number of the smaller research clusters that appeared in our earlier maps no longer make a showing in the five year map. Two examples of this are the dyslexia group that appeared in our 1985 map, and the attrition group that appeared in our 1984 map. Neither group has a presence in the 1982-86 overview map. In the case of the dyslexia group, my feeling is that this idea was not being seriously pursued by 1986. Attrition research is active in the 1982-86 period, but it has not yet developed into a significant research theme. Likewise, the *Français Fondamental* group that has appeared in some of our single year maps appears here only as a minor influence (Gougenheim) in Cluster 1,and the inlfuence of this kind of research appears to be on the wane. On the other hand, Reading in an L2 and Dictionaries and Guessing Behaviour have consolidated by 1986, and are sufficiently coherent to appear as co-citation clusters that are characteristically different from other strands of L2 vocabulary research. A notable omission from the 1982-1986 map is any sign of the ESF work that we identified as a major new growth area in 1986. This highlights a particular problem with larger window analyses in that they tend to favour research that was active at the start of the window, and are not good at picking up newer research themes.

The most surprising feature of the 1982-1986 overview map is the emergence of Krashen as by far the Most Significant Influence in this data set, and unusually strongly co-cited alongside both the applied linguists and the psychologists. My first reaction to this feature is that it was basically an artefact due to the fact that some of Krashen's earlier publications which dealt with neurolinguistics, and the implications of brain structure, were being cited by psychologists, while his more recent work, which dealt specifically with L2 acquisition, the effects of age on language learning and so on were being cited by applied linguists. In fact, this analysis is very superficial, and does not capture the complexity of Krashen's citations in this data set. Alongside the specific citations of his work, Krashen seems to be the go-to person for anyone who wants to cite a general source on second language acquisition at this time. Significantly, perhaps, the number of papers citing Krashen goes up markedly over the 1982-86 period – two in 1982, three in 1983, four in 1984, eight in 1985 and 13 in 1986. These citations mostly refer to the six(!) books that appeared in this period (*Principles and Practice in Second Language Acquisition* and *Second Language Acquisition and Second Language Learning* both published in 1982, *The natural approach: Language Acquisition in the Classroom* co-authored with Terrell in 1985, the two edited volumes co-authored with Robin Scarcella and *Language Two* co-authored with Dulay and Burt. Ironically, very little of this work deals explicitly with vocabulary acquisition – indeed, Laufer (1986) notes that *Research in Second Language Acquisition* (1980) "includes papers on communicative competence, prosodic development and syntactic development. But no vocabulary development" (p69), while "Language Two … one of the most comprehensive texts on second language acquisition does not deal with vocabulary, as if vocabulary was not part of second language acquisition"(p69).

Krashen's seminal paper *We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis,* and the much cited *Clockwork Orange* paper co-authored with Pitts and White, both published in 1989, still lie in the future. However, given the sheer volume of

Krashen's publications, we can expect him to remain a dominating influence in vocabulary research for many years to come.

5. Conclusions

This paper has provided an overview of the L2 vocabulary research that was published in a five-year window between 1982 and 1986. The bibliometric maps that we have analysed present a picture which will be recognisable to most people who were around at the time, but will perhaps be unfamiliar to younger researchers. Indeed, the maps discussed in this paper are strikingly different from maps that emerge from the more modern research. Few of the very significant modern vocabulary researchers figure in the 1982-86 map, and most of the themes which characterise current L2 vocabulary research have yet to emerge as coherent research clusters. Clearly, there is a lot of change to come, and we will be able to map these changes in future studies. In the meantime, what we have established here is a baseline against which we can evaluate new developments and paradigm shifts within L2 vocabulary research.

References

Bogaards, P. (1980). Geheugen en woordverwerving in een vreemde taal. [Memory and the acquisition of words in a foreign language.] *Levende Talen*351, 272-83.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Dulay, H., M. Burt and S. D. Krashen. (1982). *Language Two.* Oxford: Oxford University Press.

Kirsner, K., H. Brown, S. Abrol, N. Chadha and N. Sharma (1980). Bilingualism and lexical representation. *Quarterly Journal of Experimental Psychology,* 32, 585-594.

Kirsner, K., M. Smith, R. S. Lockhart, H. L. King and M. Jain. (1984). The bilingual lexicon: language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behaviour,* 23, 519-529.

Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning.* Prentice-Hall International.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition.* Prentice-Hall International.

Krashen, S. D. (1985). *The Input Hypothesis.* Harlow: Longman.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: additional evidence for the Input hypothesis. *Modern Language Journal, 73, 440-464.*

Krashen, S. D. and R. Scarcella. (1982). *Child-adult differences in Second Language Acquisition.* Rowley, Mass.: Newbury House.

Krashen, S. D. and T Terrell. (1983) *The natural approach: Language Acquisition in the Classroom.* San Francisco: The Alemany Press.

Laufer, B. (1986) Possible changes in attitude towards vocabulary acquisition research. *IRAL,* 24,1, 69-75.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences,* 16(12), 317–324.

Meara, P. M. (2012). The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal*, 43(1), 7-22.

Meara, P. M. (2014). Life before Nation: Bibliometrics and L2 vocabulary studies in 1982. In: M. Gómez González, F. J. Ruiz de Mendoza Ibañez, F Gonzálvez García & A. Downing (Eds.), *The Functional Perspective on Language and Discourse: Applications and implications.* (pp. 111-129). Amsterdam: John Benjamins.

Meara, P. M. (2015). Vocabulary research in 1983: A bibliometric analysis. *Linguistics Beyond and Within,* 1, 187-198.

Meara, P. M. (2016). Two steps backwards: A bibliometric analysis of L2 vocabulary research in 1984. *Linguistics Beyond and Within*, 2, 139-152.

Meara, P. M. (2017). A new beginning? A bibliometric analysis of L2 vocabulary research in 1985. *Linguistics Beyond and Within*, 3, 136-154.

Meara, P. M. *VARGA: the Vocabulary Acquisition Research Group Archive.* Available at: *http://www.lognostics.co.uk/varga/* (accessed 27th November 2018).

Pitts, M., H. White and S. D. Krashen. (1989). Acquiring second language vocabulary through reading: a replication of the Clockwork Orange study using second language acquirers. *Reading in a Second Language,* 5,2, 271-275.

Price, D. (1965). Networks of scientific papers. *Science*, 149(3683), 510-515.

Scarcella, R and SD Krashen (eds.) (1980). *Research in Second Language Acquisition*. Rowley, Mass: Newbury House.

Sciarone, A. G. (1979). *Woordjes leren in het vreemdetalenonderwijs*. Muiderberg: Coutinho b. v.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265-269.

Smith, M. and K. Kirsner. (1982). Language and orthography as irrelevant features in colour-word and picture-word Stroop interference. *Quarterly Journal of Experimental Psychology,* 34A, 153-170.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science,* 32, 163-171.

Wilkins, D.A. (1972). Linguistics in Language Teaching. London: Edward Arnold.

Appendix 1: Co-citation analysis: The methodology

The co-citation method was developed by Small in a number of papers published in the 1970s (e.g. Small, 1973). This approach, which was actually built on earlier bibliometric work by Price (1965), has been extensively used to analyse research in the natural sciences (e.g. White & Griffith, 1981) but does not seem to have been adopted as a standard tool by researchers in the Humanities.

The raw data for a co-citation analysis consists of a list of all the authors cited in the set of papers to be analysed. For each paper in the data set, we make a list of every author that the paper cites; for each paper, each cited author counts only once, regardless of how many times they are cited in the paper; and for a cited paper with multiple authors, each of the contributors is added to the author list.

This raw data is then used to construct a large matrix showing which authors are cited together in each of the papers in the data set. The matrix can then be analysed using a program such as **Gephi** (Bastian, Heymann, & Jacomy, 2009). Gephi performs a cluster analysis on the data, groups together authors who tend to be cited alongside each other in a number of papers, and outputs a map which shows the composition of the clusters and the relationship between them. The clusters are generally taken to represent "invisible colleges" in the data – i.e. groups of researchers who share similar reference points and a common research focus.

Appendix 2 omitted items (theses, books, unpublished items and unobtainable items)

Theses

Bahat, E

The acquisition of word formation devices in a second language. Unpublished master's thesis, Tel Aviv University. 1986.

Kruse, H
A computer word-association test as a test of second language proficiency. Unpublished master's thesis, University of Utrecht. 1986.
Quigley, JR
A semantic field approach to passive vocabulary acquisition for advanced second language learners. PhD. North Texas State University. 1986.
Verkaik, P and P van der Wijst
Taal verlies en woordherkenning in het Frans als vreemde taal. [Language loss and word recognition in French as a foreign language.] Masters Thesis. Katholieke Universiteit, Nijmegen. 1986.

Books

Crow, J
The Keyword approach: vocabulary for advanced reading comprehension. Englewood Cliffs, NJ.: Prentice Hall Regents. 1986.
Gairns, R and S Redman
Working with words: a guide to teaching and learning vocabulary. Cambridge: Cambridge University Press. 1986.
Linnarud, M
Lexis in composition: a performance analysis of Swedish learners' written English. Malmö: Liber Förlag. 1986.
Morgan, J and M Rinvolucri
Vocabulary. Oxford: Oxford University Press. 1986.
Neuner, G and E Schade
Lernschwierigkeiten ausländischer Schüler mit dem Wortschatz der Grundschule. Kassel: Kasseler Werkstattberichte zur Didaktik. 1986.

Unobtainable items

Berwald, JP
Au courant: Teaching French vocabulary and culture using the mass media. Language in Education: Theory and Practice 65. ERIC ED267644. 1986.
Cheung, YS and PLM Lee
A study of the English vocabulary in junior secondary textbooks in Hong Kong. Educational Research Establishment, Education Department, 1986.
Hatakka, JJM
We all make mistakes. *Kielikeskusuuntisia Language Centre News* (Jyväskylä) 4(1986), 25-30.
Hubbard, P, J Coady, J Graney, K Mokhtari and J Magoto
Report on a pilot study of the relationship of high frequency vocabulary knowledge and reading proficiency in ESL readers. *Ohio University Working Papers in Linguistics and Language Teaching* 8(1986), 48-57.
Löschmann, M
Die Arbeit an lexikalischen Kenntnissen. In: **G Desselmann and H Hellmich** (eds.) *Didaktik des Fremdsprachenunterrichts (Deutsch als Fremdsprache)*. Leipzig. 1986. 141–166.
Maul, S
Falsaj amikoj. *Esperanto aktuell: Organ des Deutschen Esperanto-Bundes* 5,8(1986), 5-6.
van Weeren, J
Falsche Freunde. *Grazer linguistische Studien* 25(1986), 341-342.

Unpublished items

Black, A
The effects on comprehension and memory of providing different types of defining information for new vocabulary. Cambridge: MRC Applied Psychology Unit. 1986.

LingBaW

Linguistics Beyond And Within 4 (2018), 129-140

The concept of 'touch' in the formation of the Croatian and Turkish lexicon: The example of tactile verbs

Ida Raffaelli and Barbara Kerovec
University of Zagreb, Croatia

Abstract

This paper explores the importance of the concept of 'touch' in the formation of the Croatian and Turkish lexicon. The main goals of the paper are 1) to investigate differences and similarities in conceptual mappings based on the concept of 'touch' in two typologically different and genetically unrelated languages by analyzing verbs referring to touch in Croatian and Turkish 2) to see to what extent the formation of tactile verbs differs with respect to lexicalization patterns in the two languages.

Keywords: the concept of 'touch', tactile verbs, lexicalization patterns, Croatian, Turkish

1. Introduction

The paper explores the importance of one of five sensory concepts – the concept of 'touch' – in the formation of the Croatian and Turkish lexicon.[1] Previous works dealing with sensory domains in language (Williams, 1976; Viberg, 1984; Sweetser, 1990) have shown that a) concepts from one sensory domain can be used in conceptualizing concepts from other sensory domain(s), and that b) concepts from various abstract domains are frequently and regularly conceptualized on the basis of concepts from sensory domains. According to the embodiment hypothesis within the Cognitive Linguistic theoretical framework (CL), an extension of meaning from a concrete to an abstract domain is explained via *conceptual metaphor* (cf. Lakoff & Johnson, 1980; Kövecses, 1986; Lakoff, 1987; Johnson, 1987; Kövecses, 2002). Conceptual metaphors involve understanding experiences from one conceptual domain in terms of experiences from another domain. Since our earliest and primary experience of the world is

[1] A very detailed analysis of lexicalization patterns related to the vocabulary of 'taste' in Croatian and Turkish is presented in the work of Raffaelli and Kerovec (2017). The analysis of the touch vocabulary is closely related to the previous one according to the theoretical framework and the methodology used. The reason for implementing the same framework is to collect language data that could be comparable in terms of common theoretical tenets and the same analytical apparatus.

achieved through sensory perception, perceptual experience often appears to be the basis for understanding other, more abstract phenomena. Therefore, it can be stated that conceptual metaphors are usually perceptually based, which could be the reason why there are some cross-linguistic regularities in the way we use such metaphors in linguistic expressions. However, some of the studies done so far have shown that, besides regularities, there are also some cross-linguistic differences in the way lexemes and lexical structures of perception vocabulary extend their meanings onto abstract domains, in spite of the fact that sense modalities are biologically common to all humans. Differences have been observed with regards to different language families and cultures. The main goals of the paper are 1) to investigate regularities and specificities in conceptual mappings based on the concept of 'touch' in two typologically and genetically different languages and 2) to see to what extent the formation of the touch vocabulary differs with respect to lexicalization patterns in the two languages. The term *touch vocabulary* refers to a structure of lexemes which can have very different meanings but whose roots are related primarily to the concept of 'touch'. The research in the paper is limited only to verbs. A comparative analysis of the two typologically and genetically unrelated languages, such as Croatian and Turkish, could point to differences and similarities in lexicalization processes operative in the formation of vocabularies related to the concept of 'touch'.

In the following chapters our aim is to: (i) give a brief overview of the analysis of the 'touch' domain in other linguistic approaches; (ii) present some basic theoretical and methodological tenets of the analysis, and (iii) give an analysis of the Croatian and Turkish verbs related to the concept of 'touch'. We will end the paper with some concluding remarks.

2. The concept of 'touch' in linguistic analysis

One of the most influential works in the field of meaning extensions of the lexemes related to perception is certainly the work of Eve Sweetser (1990) who has pointed to the intricate relation between the concept of 'touch' and the domain of 'emotion' (for instance, Celtic and Germanic show general homonymy in these two areas; a good example is Engl. *feel*) (Sweetser, 1990, p. 37). As one of the strongest motivating factors for this interrelatedness Sweetser notes (citing Kurath, 1921) that, as she puts it, "there is no a simple and tidy way to divide physical perception from emotions": "physical pain is bound to make the subject unhappy emotionally"; on the other hand, "physical pleasure or well-being certainly promotes a cheerful emotional state" (Sweetser, 1990, p. 44). It means that physical pain is closely related to unhappiness and vice versa, physical pleasure with happiness. The modality of touch exhibits a property that is not characteristic for other hierarchically higher modalities[2], like sight and hearing. Information gathered by the modality of touch is mostly very specific, intimate and cannot be generalized. This is the reason why the modality of touch does not conceptually relate to knowledge, but mostly to emotions. Sweetser (1990, p. 44) illustrates this claim with the story of the blind men and the elephant, pointing out that the story captures in a nutshell the reason why there is no

[2] Viberg (1984) presents a hierarchy of sense modalities and, accordingly, a hierarchy of the perception vocabulary. The touch modality is in the centre of the hierarchy, while sight is on the top and taste on the bottom of the hierarchy.

conceptual mapping from touch to knowledge, which is regular and frequent for the sight vocabulary. The only expression in Croatian that relates to knowledge by using the tactile verb is *ispipati situaciju* (lit. 'to touch a situation all over repeatedly and thoroughly') 'to investigate the situation', while in Turkish none of the tactile verbs can have a meaning related to the domain of knowledge.

A detailed cross-linguistic analysis of meaning extensions of verbs referring to touch has been conducted by Ibarretxe–Antuñano (2006). In this research on basic/generic tactile verbs in Basque, English and Spanish, Ibarretxe–Antuñano has elaborated mappings from the domain of tactile perception onto other domains of experience in a more detailed way than Sweetser has. She has identified four basic meaning extensions of tactile verbs:

1) 'to affect (physically or metaphorically)'
   e.g.: physically: *Just don't touch anything in my room.* ('change of location', 'change'), metaphorically: *John touched Mary's heart.* ('change of emotions or state')
2) 'to partake food/drink', e.g. *John hardly touched the food.*
3) 'to reach'
   spatial end-point: *El barco tocó puerto ayer.* (lit. The ship touched the port yesterday) 'The ship arrived yesterday.'
   temporal end- point: *Tocan a pagar.* 'It is time to pay.'
   touch.3.p to pay
   metaphorical end-point: *He touched the high point in his career.*
4) 'to deal with (superficially)'
   e.g. *I wouldn't touch that business.*
   He barely touched on the incident in his speech.

It can be seen that the first of these meaning extensions – 'to affect' – is linked to a wider domain which subsumes not only the domain of 'emotion' (as identified by Sweetsers) but some other domains as well, like 'change of location'. In addition, the domain of 'reaching' also includes several very different subdomains (spatial, temporal and metaphorical in the strict sense).

In this paper we have taken the results of Ibarretxe–Antuñano as a starting point for the analysis of conceptual mappings since the data of Croatian and Turkish tactile verbs show similar trends in meaning extensions. We particularly agree with the thesis that the domain of 'affecting' is more appropriate than just the domain of emotions since it is an overarching structure for a variety of experience domains. However, typological differences between the two languages we have dealt with required a more comprehensive analytic approach of the tactile verbs.

One of the main typological differences between Croatian and Turkish is the existence of prefixation in Croatian as one of the main word-formation processes enabling the formation of new verbs. Therefore, one of the aims of this research is to define the way prefixation – which does not exist in Turkish – can influence the formation of the lexicon. In other words, the goal is to investigate how this word-formation process is related to the lexicalization of experiences

from different domains and what are the differences and similarities with Turkish as an agglutinative language, in which suffixation is a dominant word-formation process.

As we will show further in this paper, such an approach enables: 1) the investigation of morphological (derivational) devices used in building lexicon and influencing the change of meaning and 2) a more fine-grained classification of the meaning extensions (and conceptually related domains), 3) an insight into similarities and differences in meaning extensions driven by grammatical differences between prefixation in Croatian and agglutination in Turkish.

3. Theoretical framework and methodology

The model that we find suitable for pointing to differences and similarities in lexicalization of the concept of 'touch' is the model of lexicalization patterns[3]. The term *lexicalization pattern,* as used in this paper, is closely related to Talmy's (1985) notion of lexicalization patterns.[4] There are, however, some differences between our approach and Talmy's. The main difference is that we define lexicalization patterns mostly with respect to the stem or root that has served as a basis for the formation of newly formed (derived) lexemes. Thus, the lexemes under investigation are morphologically and semantically related. Similar to Talmy, we believe that the analysis of lexicalization patterns consists in defining diverse lexical forms that capture the way the world is conceptualized. Thus, it is necessary to define types of lexicalization patterns that lexicalize certain concepts or conceptual mappings within a certain language or among languages. The model of lexicalization patterns defines morphological and semantic features that are similar or different for lexemes that share the same formal (stem) or conceptual basis. Therefore, it is possible to recognize conceptual and semantic extensions captured by the same or different lexicalization patterns – which is not the case in Talmy's model. Moreover, the model of lexicalization patterns (as developed by Raffaelli & Kerovec, 2017) shares an important feature with Talmy's model: the notion of typological relevance. Some lexicalization patterns can be language specific, and some can be regular and frequent among languages, mostly within a specific language family.

The typological relevance of the model of lexicalization patterns is foremost seen in the ability of the model to define diverse strategies used in building lexicon as well as lexical forms that capture the way the world is conceptualized in genetically close or remote languages. The analysis of the language data will point to three different typological features that influence lexical structures in the two languages.

---

[3] The model has already been implemented in the analysis of the vocabulary related to the concept of 'taste' in Croatian and Turkish. See Raffaelli and Kerovec (2017).

[4] Talmy (1985, p. 57) defines lexicalization patterns as a relation between meaning and surface expressions and investigates which semantic elements are expressed by which surface elements. Semantic elements of different types can be expressed by the same type of surface element, as well as the same type of semantic element can be expressed by several different surface elements. A range of typological patterns and universal principles can be found by such an approach.

3.1. Methodology

In the analysis that follows, verbs primarily related to the concept of 'touch' have been selected and each of them has been checked for meanings and contextual uses in Croatian and Turkish dictionaries[5] and in the Croatian National Corpus (CNC), Croatian Web Corpus (HrWaC) and METU Turkish Corpus. Corpus-based data enabled to point to a regular and frequent usage of the analysed verbs. Croatian language is taken as a starting point in the analysis and Turkish vocabulary is compared with respect to lexicalization patterns established for Croatian.

In the next chapters we will try to point to:

1) the main typological differences concerning a diversity of lexicalization patterns in the two languages which influence meaning extensions;
2) the main regularities and specificities in meaning extensions (conceptual mappings) of the tactile verbs in the two languages.

4. Croatian and Turkish verbs referring to the concept of 'touch' and their semantic extensions

According to the corpus and dictionary-based data, in Croatian there are three, and in Turkish two main verbal roots related primarily to the concept of 'touch', which serve as a basis for the formation of tactile verbs. In Croatian these are *tak, dir* and *pip* and in Turkish *dokun* and *değ*.

4.1. Aspect as a typologically relevant feature

The first typological difference between Croatian and Turkish is that in Croatian the category of verbal aspect is expressed morphologically while in Turkish this is not the case. Thus, three main Croatian tactile verbs have their perfective and imperfective forms as follows:

***do**tak**nu**ti, tak**nu**ti*, perf./ *tic**a**ti*,[6] imperf. 'to touch'
***do**dir**nu**ti, dir**nu**ti*, perf. / *dir**a**ti* imperf. 'to touch'
popipati, pip**nu**ti perf. / pip**a**ti imperf. 'to touch'

On the other hand, two main Turkish verbs related to touch do not change morphologically to express aspect (aspect is denoted by verb tenses and periphrastic constructions):

dokunmak (perf. and imperf.) 'to touch'
değmek (perf. and imperf.) 'to touch'

This is an important difference between the two languages because in Croatian the morphological change affects not only the change of aspect, but additionally, different aspectual forms have different concrete and abstract meanings, i.e. they are semantically related to

[5] For the dictionaries consulted in the research see the list at the end of the paper.

[6] It has to be pointed out that the aspectual pair *taknuti*, perf. and *ticati*, impf. do not share the same etyom, i.e. their origins are diverse with respect to Old Slavic lexemes. However, they have become close over time and are nowadays considered as an aspectual pair as in ***do**taknuti* – ***do**ticati* 'to touch superficially'.

different experiential domains. For instance, the verb *taknuti*, perf. extends its meaning to the domain of emotions (e.g. *taknuti koga u srce* 'to touch someone's heart', or *taknuti u živac* 'to touch someone's nerve', 'to provoke') while the verb *ticati*, impf. is linked to the domain of 'concern' or 'relatedness' (e.g. *To se tiče svakoga od vas.* 'It concerns all of you / it is related to all of you'). Moreover, there is a systematic difference between *taknuti* and *ticati*. While *taknuti* can refer to the concrete domain of touch as well as to the domain of emotions, *ticati* cannot. Its only meaning is 'to concern'. The same holds for *dirnuti* 'to move emotionally', *dirati* 'to bother' (e.g. *To me uopće ne dira* 'It doesn't bother me at all'), and *pipnuti* (in negative constructions such as *Nisam ni pipnula juhu/knjigu.*[7] in a broader meaning 'not to consume at all', 'I didn't even taste the soup / I didn't even open the book'. All the verbs, except the verb *ticati*, refer to the domain of touch and to one or more abstract domains. Additionally, the Croatian suffix *-nu-*, that appears in the verbs *tak**nu**ti, pip**nu**ti* and *dir**nu**ti* (all perfective verbs) carries a diminutive meaning, i.e. the meaning that refers to the activity being performed in a small quantity, superficially and at once. As pointed out by Katunar (2013: 9) there is a difference between the diminutive suffix *-nu-* and *-k-* as in *pip**k**ati, dir**k**ati*. *Pipnuti* and *dirnuti* are perfective forms referring to an activity that has happened superficially and rapidly only once, whereas verbs formed with the suffix *-k- pipkati* and *dirkati* are imperfective verbs and refer to a repeated action of touching somebody or something as in: *Nježno ju je dirkao laticama cvijeta.* 'He touched her gently with flower petals.' In Turkish, all these meanings of different forms of Croatian verbs can be conveyed by the same form of the verb *dokunmak* (except the meaning of 'concern' like in *To se tiče svakoga od vas.* 'It concerns all of you' which cannot be expressed by Turkish tactile verbs): *Gözyaşları bana dokundu.* 'Her tears touched me (emotionally).', yumurtaya bile dokunmayacak *kadar vejetaryen* 'vegetarian to the degree that he wouldn't even touch (consume) an egg', *birinin sinirine dokunmak* 'to get on one's nerves' (lit. to touch one's nerve), *eğitim konusuna dokunan bir yazı* 'a document related to the question of education'. The meaning extension of tactile verbs to the domain of 'bothering' or 'disturbing' can be seen in both languages, but in Turkish the verb *dokunmak* additionally extends its meaning to the domain of harming health, again without any morphological change: *Bu hava/yemek bana dokunur.* 'This weather/food harms me (harms my health)' (lit. 'This weather/food touches me.'). As far as repeated action is concerned, in Turkish it can be expressed by a duplicated form of the converb *–(y)A* signaling the intensity of an action (e.g. *karanlıkta dokuna dokuna* 'tapping in the dark'). The diminutive meaning is again not expressed morphologically, but lexically, either by 1) an adverb signaling diminutive meaning such as *hafifçe* 'gently'+ *dokunmak* 'gently touch', or 2) the verb *değmek* 'to touch (slightly)'. The verb *değmek* is, in comparison to the verb *dokunmak,* more likely to be used in contexts where the meaning of superficial contact should be conveyed: *Kapıdan bir an birbirimiz**e** değerek girdik.* 'We entered through the door touching one another for a moment'.

[7] It has been suggested (Barcelona, p.c.) that instead of having the meaning 'to partake of food or drink', which is too specific, it would be better to propose a more general meaning like 'to partake of something'. While Ibarretxe–Antuñano thinks that it should be reserved only to food and drink, our example shows that it can be extended to other domains as well.

Turkish tactile verbs *değmek* and *dokunmak* extend their meanings to the domain of 'reaching' more productively than Croatian tactile verbs in spite of the fact that they do not change morphologically: e.g. *Mektup elim**e** değdi.* 'I received the letter' (lit. 'The letter touched my hand'), *Kurşun hedef**e** değdi.* 'The bullet hit the target' (lit. 'The bullet touched the target'). The meaning of 'reaching' of the verb *değmek* extends additionally to the abstract domain of 'value', in the sense of reaching the price of something, e.g. *Ömrümün üç günü üç yüz bin liraya değer.* 'Three days of my life are worth three hundered lira.', lit. 'Three days of my life touch (reach) three hundered lira.'. The fact that we must reach something in order to touch it triggered a semantic extension of the verb *dokunmak* too, but its extended meaning is related more to the abstract domain of 'affecting', as in *Onun benim gibi çok kişiye yardımı dokundu.* 'He helped many people like me', lit. 'His help touched (reached) many people like me', *Felsefenin bana çok* faydası dokundu. *'I have benefited a lot from philosophy.', lit.* 'A lot of philosophy's benefit touched (reached) me.'

4.2. Prefixation as a typologically relevant feature

The second important typological difference between Croatian and Turkish is prefixation which exists in Croatian but not in Turkish as an agglutinative language. Prefixes in Croatian can trigger divers meaning changes. There is a significant difference between *taknuti* vs. ***do**taknuti* and *dirnuti* vs. ***do**dirnuti*. *Dodirnuti* and *dotaknuti* lexicalize a less superficial activity, with an implication that the activity has lasted for a certain period of time. This is especially visible in metaphorical usage, as in: *Nakon toga njihov je odnos dotaknuo dno.* 'After that their relationship has reached the bottom'.[8] In this expression, a prior event to the one lexicalized by the expression is implicated. In other words, something happened before and caused the final stage of their relationship.

Another lexical refinement in the physical domain can be observed for the verb *taknuti*: different prefixes contribute to the modification of spatial meanings, which can be considered as meaning extensions from the domain of touch to the spatial domain in the strict sense:

***na**taknuti* 'to put/slip/fix **on**'
***za**taknuti* 'to hook **on**', 'to stick/push/thrust/tuck **in**'
***u**taknuti* 'to push/thrust/fix/wedge **in(to)**', 'plug up, socket'

Such a lexical refinement of the physical features of the touch modality in Croatian cannot be expressed by Turkish tactile verbs.

Furthermore, some prefixed forms of Croatian tactile verbs convey metaphorical, abstract meanings which cannot be conveyed by Turkish tactile verbs. One of the verbs which is regularly used in meanings related to abstract domains is the verb *dotaknuti*. The verb ***do**taknuti* is very frequently used in the meaning 'to gently affect' or 'to mention' as in: *To je zemlja koju je turizam tek dotaknuo.* 'This is a country that has only recently been affected/touched by tourism.' or *U svom je govoru dotakla mnoga aktualna pitanja.* 'She mentioned / touched upon

[8] This is the only expression in which the Croatian tactile verb is used in the meaning of 'reaching'.

many current issues in her speech.'. The verb ***po****taknuti* has exclusively metaphorical meaning 'to stimulate', 'to encourage', 'to animate' and ***is****taknuti* means 'to stress/emphasize', 'to point out', 'to bring into prominence'. As can be seen, *dotaknuti, potaknuti* and *istaknuti* share the same verbal stem, i.e. they are derived from the verb *taknuti*. However, they differ with respect to their semantic structures. *Dotaknuti* can refer to touch as well as to abstract domain of 'affecting' and 'mentioning' whereas *potaknuti* and *istaknuti* cannot refer to touch. They convey exclusively abstract meanings. It must be pointed out that Turkish tactile verbs cannot express the meanings 'to stress/emphasize'and 'to stimulate', 'to encourage', 'to animate'. As for the meaning 'to mention', it can be conveyed by the reflexive form of the verb *değmek*, which will be analysed in the subchapter "Suffixation as a typologically relevant feature".

4.3. Prepositions as a typologically relevant feature

The third typological difference is related to prepositions which do not exist in Turkish. In Croatian different preposition when combined with a tactile verb influence the change of its meaning, e.g. with the preposition *po* 'along/over' the verb *dirnuti* has exclusively concrete meaning (*dirnuti* ***po*** *ramenu* 'touch (upon) one's shoulder') while with the preposition *u* 'in' conveys the meaning related to emotions (*dirnuti* ***u*** *srce* 'to touch one's heart'). It is interesting that the preposition *u* 'in', when combined with the verb *taknuti*, represents a construction[9] – *taknuti u* – which conveys a very specific meaning 'to affect something or somebody deeply'. Namely, *taknuti* as a diminutive verb that refers to a superficial and brief action, when combined with the preposition *u* as in *taknuti u* denotes an abstract event that has a severe and profound impact on someone or something: e.g. *taknuti u dušu* 'to touch someone's soul', *taknuti u živac* 'to irritate someone', *taknuti u tradicijske vrijednosti* 'to touch traditional values'.

Although Turkish postpositions can in some aspects be compared to Croatian prepositions, they do not combine with tactile verbs; Turkish verbs *dokunmak* and *değmek* always combine with the same case, no matter whether a meaning is physical (concrete) or metaphorical (abstract), and that case is dative, morphologically denoted by the suffix *–(y)A*, e.g.:

*omzun**a** dokunmak* 'to touch (upon) one's shoulder'(physical touch)
*kalbin**e** dokunmak* 'to touch one's heart' (emotions)
*Elim onunkin**e** değdi.* 'My hand slightly touched his (hand)' (physical touch)
*Onun için her şeyi feda etme**ye** değer* 'For her it's worth to sacrify everything' (value)

4.4. Suffixation as a typologically relevant feature

The fourth typological difference is related to factitive/causative and reciprocal categories which are in Turkish expressed by different suffixes. These suffixes, when added to the root *değ-* (referring primarily to the concept of touch) trigger a semantic extension to the domain of

[9] The role of governed prepositions in the formation of the V PP constructions in building lexicon is comprehensively analysed and described in Katunar (2015).

'affecting', but to a subdomain which cannot be lexicalized by Croatian tactile verbs, and that is the subdomain of 'change' (in the strict sense):

değmek 'to touch' → *değ-**iş**-mek* 'to change (intrans.)' → *değ-**iş**-**tir**-mek* 'to change (trans.)'

These meanings can be subsumed under the broader domain of 'affecting'. The motivation for this meaning extension could be explained by the fact that in order to change something, we must touch it.

On the other hand, when the reflexive suffix *–in* is added to the same root *değ-*, it changes its meaning in a similar way the Croatian reflexive pronoun *se* does in combination with the prefixed form of the diminutive verb *tak**nu**ti*:

*değ-**in**-mek* vs. ***do**tak**nu**ti **se*** 'to deal with / to comment on / to talk about / to mention'
e.g. Tur. *kimlik konusuna değ**in**mek* vs. Cro. *dotaknuti **se** teme identiteta* 'to mention / to talk about the matter of identity'

5. Discussion

All the data analyzed so far show some semantic specificities and regularities between Croatian and Turkish with respect to diverse lexicalization patterns. Finally, it is necessary to compare differences and similarities between Croatian and Turkish in lexicalization of conceptual domains and to shed light on some conceptual subdomains within the domains already defined by Ibarretxe–Antuñano (2006). The table (1) below shows that in this respect overlaps between the two languages are only partial. Even in the physical domain of 'touching', Croatian shows a more fine-grained lexicalization of diverse physical activities. For instance:

1) The verbs *taknuti, pipnuti* and *dirnuti* lexicalize a punctual event – something that happened very rapidly, superficially and once. Both verbs are derived via a diminutive suffix *-nu-* referring to the activity that is performed in a small quantity or intensity. On the other hand, verbs *pipkati, dirkati* are also diminutive verbs, formed with the suffix *-k-*, but referring to a repeated, successive action. In Turkish such a difference cannot be lexicalized by a morphological change of a verb.
2) The imperfective verb *pipati* 'touch, palpate', which does not have its simple lexical equivalent in Turkish, expresses a durative activity. It differs from *dirati* (also imperfective form) in the sense that it lexicalizes (very often) a chain of tactile movements that we perform in order to find something (*Pipala sam po mraku zidove da nađem zvono na vratima* 'I've been touching the walls in the dark to find the doorbell'). In the fourth row of the table it can be seen that the meaning 'to find something by touching' in Croatian (but not in Turkish) extends to the domain of knowledge: the prefixed verb ***is**pipati* conveys the abstract meaning 'to investigate/inquire' with the result of revealing something, as in already mentioned example *ispipati situaciju* 'to investigate a situation'.
3) Different spatial meanings of fixing something *on* or *in(to)* and which are conveyed by prefixed forms of the verb *taknuti* (***na**taknuti, **za**taknuti, **u**taknuti*) cannot be lexicalized by Turkish tactile verbs.

4) The meaning of 'physically reaching something' can be expressed by a tactile verb only in Turkish (*Mektup elim**e** değdi* lit. 'the letter touched my hand' *I received a letter*). Additionally, Turkish verbs *değmek* and *dokunmak* are more productive when meaning extensions via the domain of 'reaching' are concerned.
5) In the fifth row of the table it can be seen that the domain of 'affecting' includes several different, but conceptually related subdomains which are differently lexicalized in the two languages. The subdomains of 'affecting emotionally' and 'disturbing', 'bothering' overlap, but in Croatian the verbs change morphologically while in Turkish they do not (and this tendency can be observed in Turkish in all the domains lexicalized by tactile verbs, which can be seen in the last right column of the table). The meanings of 'disturbing health' and 'change' are lexicalized only in Turkish, while the meaning of animating, initiating or encouraging only in Croatian.
6) The subdomain of 'dealing with' or 'concerning' is lexicalized in both languages, while the meanings 'to emphasize' (Croatian) and 'to be worth' (Turkish) are not.

Table 1: *Lexicalization of diverse domains and subdomains*

		CROATIAN	**TURKISH**
PHYSICAL TOUCH		*tak**nu**ti, dir**nu**ti* – punctual, superficial	DAT + *değmek*
		pipati – chain of tactile movements	DAT + *dokunmak*
FIXING (ON/IN)		***na**taknuti* vs. ***za**taknuti* vs. ***u**taknuti*	**X**
REACHING (→Tur. affecting)		***do**taknuti dno* 'to reach the bottom' (abstract)	DAT + *değmek*
FINDING (→ knowledge)		***na**pipati* 'to feel / find by touching'	**X**
		***is**pipati* 'to investigate/inquire' (abstract)	
AFFECTING	**Emotions**	*dirati **u**, dirnuti **u***	DAT + *dokunmak*
	Disturbing / bothering	*dirati* (+ACC)	DAT + *dokunmak*
		***za**dir**k**ivati* 'to tease' (+ACC)	
	Disturbing health	**X**	DAT + *dokunmak*
	Benefiting	**X**	DAT + *dokunmak*
	Initiate / encourage	***Po**taknuti*	**X**
	Change (general)	**X**	*değ**iş**mek* (intrans.)
			*değ**iş**tir**mek* (trans.)
DEALING WITH (concern)		*ticati se*	DAT + *dokunmak*
		***do**tak**nu**ti (se)*	DAT + *değ**in**mek*
EMPHASIZING / POINTING OUT		***is**taknuti (se)*	**X**
VALUE (reaching the price)		**X**	DAT + *değmek*

6. Conclusion

A contrastive analysis of two languages such as Croatian and Turkish showed that there is a need for introducing a more comprehensive analytic approach to tactile verbs, as to any other perception vocabulary. Turkish and Croatian tactile verbs show the potential for a more fine-grained classification of conceptually related domains. As our analysis shows, Croatian verbs vary in lexicalization already in the domain of a physical sense modality due to prefixation and suffixation. Moreover, the model of lexicalization patterns points to diverse grammatical

devices that encode a certain meaning, as in the examples of V PP constructions – *taknuti po*, *taknuti u* and others. These constructions where different prepositions are used differ significantly in meanings. The construction *taknuti po* refers to the physical, whereas *taknuti u* is related to the domain of emotions.

Defining lexicalization patterns provided an in-depth semantic analysis of how certain conceptual and semantic structures are captured and conveyed by diverse grammatical devices. This kind of approach enables a more fine-grained definition of conceptual domains and subdomains that are lexicalized via morphologically and semantically close verbs in the two typologically different languages. As shown in our analysis, differences between the two languages can be seen only if subdomains are observed – as illustrated in the Table 1, Croatian and Turkish show very little overlaps on the level of subdomains. The defined domains and subdomains represent encyclopedic knowledge that is encoded in different lexical forms in the two languages.

Finally, it has to be noted that it was not possible to give an overall fully exhaustive analysis of the lexicalization patterns related to tactile verbs in the two languages due to their complexity. However, we believe that the main similarities and differences between the two languages have been pointed out. It can be concluded that the model of lexicalization patterns enables to capture typological regularities as well as specificities with respect to the diversity of grammatical devices languages use to encode certain meanings.

References

Ibarretxe-Antuñano, I. (2006). Cross-linguistic polysemy in tactile verbs. In June Luchjenbroers (Ed.), *Cognitive Linguistics Investigations: Across languages, fields and philosophical boundaries* (*Human Cognitive Processing* 15) (pp. 235-253). Amsterdam, Philadelphia: John Benjamins.

Johnson, M. (1987). *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason.* Chicago/London: The University of Chicago Press.

Katunar, D. (2013). Diminutives in Action: A acognitive account of diminutive verbs and their suffixes in Croatian. *Suvremena lingvistika, 75*, 1-24.

Katunar, D. (2015). *Ustroj leksikona u konstrukcijskoj gramatici – primjer prijedloga u hrvatskom jeziku* (Doctoral dissertation). Zagreb.

Kövecses, Z. (1986). *Metaphors of Anger, Pride, and Love: A Lexical Approach to the Structure of Concepts.* Amsterdam, Philadelphia: John Benjamins.

Kövecses, Z. (2002). *Metaphor: A Practical Introduction.* New York: Oxford University Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Raffaelli, I. (2017). Conventionalized patterns of colour naming in Croatian. In Cergol Kovačević, K., & Udier, S. L. (Eds.), *Applied Linguistics Research and Methodology. Proceedings from the 2015 CALS conference* (pp. 171-186). Franfurt am Main/Bern/Bruxelles: Peter Lang Verlag.

Raffaelli, I., & Kerovec, B. (2017). The concept of 'taste' in formation of Croatian and Turkish lexicon: A contrastive analysis. *Suvremena lingvistika*, *83*, 21-48.

Sweetser, E. (1990). *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantics.* Cambridge: Cambridge University Press.

Viberg, Å. (1984). The verbs of perception: A typological study. In Butterworth, B., Comrie, B., & Dahl, O. (Eds.), *Explanations for language universals* (pp. 123-162). Berlin: Mouton de Gruyter.

Williams, J. M. (1976). Synaesthetic Adjectives: A possible law of semantic change. *Language, 52* (2), 461-478.

Dictionaries

Damjanović, S. (2004). *Mali staroslavensko hrvatski rječnik*. Zagreb: Matica hrvatska.
Đinđić, S., Teodosijević, M., & Tanasković, D. (1997). *Türkçe-Sırpça Sözlük*. Ankara: Türk Dil Kurumu.
Hrvatski enciklopedijski rječnik. (2002). Zagreb: Novi liber
http://www.nisanyansozluk.com/
Püsküllüoğlu, A. (2005). *Arkadaş Türkçe Sözlüğü.* Ankara: Arkadaş Yayınevi.
http://tureng.com/
Türkçe Sözlük. (1983). Ankara: Türk Dil Kurumu Yayınları.

LingBaW *Linguistics Beyond And Within 4 (2018), 141-160*

Sigmatic plurals in Romance varieties spoken in Italy and their interaction with *-i* plurals

Leonardo M. Savoia, Benedetta Baldi, M. Rita Manzini
University of Florence, Italy

Abstract

In Sardinian, Friulian, Rhaeto-Romance, Occitan (not considered in this work) and Franco-Provençal varieties spoken in peripheral Italian areas, the *-i* inflection is not totally eradicated but interacts with plural *-s*. The coexistence of *-s* and *-i* reflects syntactic constraints. Specifically, *-i* is in complementary distribution with *-s* or it combines with *–s*, giving rise to a duplicated lexicalization of plurality. In any event, it is specialized for a subset of the morpho-syntactic contexts. The distribution of *-i* generally involves the D domain, i.e. determiners and clitics. In some cases, *-i* is limited to the D elements. Different agreement systems on D and on N emerge, recalling partial or asymmetric agreement phenomena known in literature. Moreover, in the relevant varieties *(-)i* is in turn the inflection of the dative clitic; this suggests that *(-)i* is endowed with a slightly different content from plural *-s*, that, on the contrary, shows no connection with dative.

Keywords: morphology, plural, dative, agreement asymmetries, Romance variation

1. Introduction

In Italian and Romanian varieties plural-*i* (and *-e*) contrast with plural *-s* in West Romania (Meyer-Lübke 1899, Lausberg 1971, Sauzet 2012). In Sardinian, Friulian, Rhaeto-Romance, Occitan (that we will not consider in this work) and Franco-Provençal varieties spoken in peripheral Italian areas, the *-i* inflection is not totally eradicated but interacts with *-s*. The compresence of *-i* and *-s* in these languages can be viewed as the reflex of the protracted competition and as the result of the old contact between the two types of number morphology.

The coexistence of *-s* and *-i*, however, is not accidental, episodic or lexicalized, but is governed by a set of syntactic constraints. More precisely, *-i* is specialized for a subset of morpho-syntactic contexts where it is in complementary distribution with *-s* or it combines with *-s* giving rise to a duplicated lexicalization of plurality. Specifically, the distribution of *-i* generally involves the D domain, i.e. determiners and clitics. In some cases, *-i* is limited to the D elements, configuring different agreement systems on D and on N, and recalling the partial or asymmetric agreement systems known in the literature (Costa and Figueiredo 2002, Bonet

et al. 2015). Moreover, in these varieties *(-)i* is the inflection of the dative clitic; this suggests that *(-)i* is endowed with a slightly different content from the plural *-s*, that, on the contrary, shows no connection with dative.

The data we discuss in this work[1] involve the connection between number, gender and agreement within DP and IP and raise the question on the nature of the inflectional elements inside N. Our model adopts the idea that morphosyntactic structures are projected from lexical items (Chomsky 1995, 2005). It assumes a morpheme-based analysis of inflectional phenomena and in general of the internal structure of the word, and that the same basic computational mechanisms underlie syntax and morphology (Manzini and Savoia 2005, 2011a,b, 2017a,b, forthcoming, Savoia et al. 2017a,b).

2. Sardinian *-s* systems

In Sardinian systems[2] the plural is expressed by the specialized *-s* morpheme. In order to understand the examples, we note that in Sardinian varieties the ending /–s/ is subject to two types of allomorphic alternation. *–s* in final position of sentence is generally followed by an epenthetic vowel copying the immediately preceding vowel, as in */feminas/* → … *feminaza*, in (2a'). Moreover, in the internal contexts of sentence the final /-s/ undergoes lenition and assimilative changes depending on the initial consonant of the following word giving rise to an interesting micro-variation (Molinu 2013, Savoia 2015). So, when *–s* precedes a voiced obstruent, it changes to *-r* e.g. in *feminar bɛttsaza* (<*feminas bɛttsaza*), in (1a') for Ardauli; in other dialects a palatal outcome occurs, for instance *-ʎ* in *femina-ʎ vɛtts-a-za* 'women old' in (3a) for Luras. In (2a') the sequence formed by the final *–s* and the initial *f-* changes to *ʃ*, e.g. *i ʃemminaza* (< *is femminaza*), in (2a') for Orroli. Finally, in (3), *-s* assimilates to the following voiced sonorant e.g. in *kussa llibbrɔzɔ?*

Consider now the internal organization of the nominal inflection. *-s* follows the nominal class vowel inflection. In Northern Sardinian, *-a-* for the 1st class, *-o-/-u* for the 2nd class and *-e-* for the 3rd class. We will concentrate on the 1st and 2nd class morphology. In the 2nd class, *-u* introduces the masculine singular and closes the noun; on the contrary, *-o-* occurs in-between the base and the plural *–s* only signalizing the masculine class; besides, *-o* closes the word in a small subset of nouns, like *ɔtt-o* 'eight'. In the feminine *–a(-)* occurs both as the singular inflection and the internal class inflection when followed by *–s*. So, we associate *–a(-)* with the f(eminine), *-u* with msg(asculine singular) and *–o-* with m(asculine). Finally, *-i* characterizes the dative.

In (1a)-(1a') we illustrate the singular and plural forms for 1st class nouns, where *-s* follows the *-a-* inflection. In the masculines of the 2nd class a split between the singular and plural vocalic inflection shows up, whereby *-u* is the ending of the singular while in the plural the intermediate vocalic specialized inflection *-o-* occurs, as evidenced by the comparison between (1b) for the

[1] All the data examined in this article have been gathered through field work with native informants in the last few years. We sincerely thank them for their collaboration.

[2] For a comprehensive view cf. Blasco Ferrer and Contini 1988, Jones 1993.

singular and (1b') for the plural. In (1c), dative clitics, singular and plural, are exemplified, where the *-i* inflection occurs.

(1) a. s-a/kuss-a vemin-a ɛtts-a
the-f/this-f woman-f old-f
'the/this old woman'

a'. s-a-s/kus-a-s femin-a-r bɛtts-a-za
the-f-pl /this-f-pl women-f-pl old-fsg-pl
'the/these old women'

b. kuss-u att-u etts-u
this-msg cat-msg old-msg
'this old cat'

b'. kuss-o-z att-o-r bɛtts-o-zɔ
this-m-pl cat-m-pl old-m-pl
'these old cats'

c. ḍ-i / ḍ-i-r dza-ða ɣust-u
him-her-dat / them-dat give-3sg this-msg
'he gives him/her/them this'

Ardauli

In Southern Sardinian dialects the feminines present the same inflection, as in (2a,a'). In the case of the masculines, the original mid post-tonic vowels have turned into high vowels owing to a neutralization process that has obscured the alternation between *-u* as msg ending and the original internal inflection *-o-* in the 2nd class masculine nouns. This distribution characterizes also object clitics, in (2c). As a consequence, there is no longer a specialized vocalic inflection for the plural and *-u(-)* occurs both in singular, in (2b) and in plural, as in (2b'). In Southern varieties the morpheme *-i* lexicalizes the dative in clitics, as in (2c'), and, moreover, it occurs in the plural determiners, indifferently in masculine and feminine contexts, as exemplified in (2a',b'). For reasons of descriptive clearness we distinguish the plural reading pl from the dative one, except arrive to a common treatment in Section 5.

(2) a. s-a vemmin-a / s oriɣ-a
the-f woman-f / the ear-F
'the woman' 'the ear'

a'. i ʃemmin-a-za (< *is femminaza*) / i-z oriɣ-a-za
the-pl women-f-pl / the-pl-pl ear-f-pl
'the women' 'the ears'

b. s-u att-u (< *su gattu*)
the-m cat-m
'the cat'

b'. i ɣatt-u-zu (< *is gattuzu*)
the-pl cat-m-pl
'the cats'

c. ḍ-u/ḍ-a/ḍ-u-zu/ḍ-a-za bbi-u
him-m/her-f/them-m-/-f-pl see-1ps
'I see him/her/them'

c'. ɖɖ-i/ɖɖ-i-zi (d)ɔna-ða ɣust-u
him/her-dat /them-dat give-3sg this-m
'He gives him/ her/ them this'

Orroli

Summarizing, in Sardinian dialects the *(-)i* inflection has the following distribution:

- It lexicalizes the dative interpretation in clitics (cf. Italian *gli*), as in (1c) and in (2c').
- In Southern Sardinian dialects the definite determiner *i-s* combines the inflection *i-* with *-s* in the masculine and feminine plural, as in (2a',b').

Some Northern Sardinian varieties (cf. Sanna 1975) present a specialized distribution of the plural vowel morphemes, giving rise to a particular type of syncretism in the plural. In the Luras dialect, in (3), the syncretism involves determiners and adjectives, which in the plural select the internal inflection *-a-* both in 1st class feminine (3a') and in the 2nd class masculine (3b'). In turn, object clitics (OCl) have only one form *l-a-s* in the plural, as in (3c). Again *–i* the dative inflection in object clitics, in (3c').

(3) a. s-a/un-a vemin-a ɛtts-a
the-f/one-f woman-f old-f
'the/an old woman'

a'. s-a-s femin-a-za / kuss-a-s femin-a-ʎ vɛtts-a-za
the-f-pl woman-f-pl / this-f-pl women-f-pl old-f-pl
'the women' / 'those old women'

b. s-u / kuss-u libbr-u s-u ðið-u
the-msg / this-msg book-msg the-msg finger-msg
'the/this book' 'the finger'

b'. s-a(-s) / kuss-a(-s) llibbr-ɔ-zɔ trɛ(-s) bbɛll-a(-s) llibbr-ɔ-zɔ s-a-ʎ dið-ɔ-zɔ
the-f-pl / those-f-pl books-m-pl three-pl nice-f-pl book-m-pl the-f-pl finger-m-pl
'the/ those books' 'three nice books' 'the fingers'

c. l-u / l-a / l-a-s[l] jam-a-na
him-msg / her-f / them-f-pl call-3pl
'they call him/her/them'

c'. l-i / l-i-(s) ða-na ɣust-u
him-her-dat / them-dat-pl give-3pl this-msg
'they give him/her/them this'

Luras

The label f for *–a-* in the examples in (3) has only a descriptive character. In fact, in this dialect it occurs both in feminine and in masculine agreement contexts, hence implying a deeper common property.

2.1. Rhaeto-Romance varieties

2.1.1. Friulian

Microvariation in Friulian brings into focus the special nature of *(-)i*. In the data in (4a-b), from *Comeglians* (Central Friuli), *(-)i* is in complementary distribution with the *-s* plural morpheme. *–s* occurs in the feminines in (4a') and in a subset of the masculines, in (4b'). In the masculine plural a threefold possibility is attested, i.e. *–i*, *-s* or a palatalized outcome of the final obstruent – e.g. *dinc* 'teeth-, as in (4b'). *i* occurs as the morpheme of the masculine plural in determiners, in (4b'), and in OCl *i-u,* in (4c); in subject clitics (SCl), in (4c), plural *(-)i* occurs in the 1st/2nd plural person and in the 3rd plural person *a-i,* where it alternates with the form *a*. Finally, *-i-* is associated with the dative clitic, as illustrated by the examples in (4d) concerning the singular dative; it is of note that plural dative has a specialized form *ur* (< Latin *illorum*), as in (4d').

(4) a. l-a / kɛ fɛmin-a vɛc-a
the-f / that woman-f old-f
'the/this old woman'

a'. l-a-s / kɛ:-s (vɛc-a-s) fɛmin-a-s (vɛc-a-s)
the-f-pl / those-pl (old-f-pl) women-f-pl (old-f-pl)
'the/those old women'

b. kɛl bjɛl ɟat / l dint
that nice cat / the tooth

b'. ke-i / i ɟat-s / i dinc / i cave-i
those-pl / the.pl cat-pl / the.pl teeth.pl / the.pl hair-pl
'those/the cats / the teeth / the hair'

ke-i bje-i / vɛc-u-s ɟats
those-pl nice-pl / old-m-pl cat-pl
'those nice/old cats'

c. i durm-iŋ / i durm-i:s / a(-i) duarm-int
SCl-pl sleep-1pl / SCl-pl sleep-2pl / SCl-pl sleep-3pl

c'. tu l-u /l-a /i-u / l-a-s iouk-s
you him-m / her-f / pl-m / her-f-pl see-2sg
'you see him/her/them'

d. a i da keʃt
SCl him/her-dat gives-3sg this
'He gives him/her this'

d'. a ur da keʃt
SCl them.pl give.3sg this
'He gives them this'

Comeglians

In other varieties of West Friuli, such as San Giorgio della Richinvelda, *(-)i* is even more scattered in the paradigm, as illustrated in detail in (5). *(-)i* occurs as the inflection of the plural in the feminines, where it is followed by the *-s* morpheme, as shown in (5a). In some sub-classes of feminines we find only *-s*. As regards masculines, two systems show up: *-s* is inserted in many nouns with the base in consonant, as in (5b), *-i* occurs in a subclass of nouns ending in lateral as in (5b'). The final *-i* that characterizes a sub-set of singular masculines, as *libr-i* 'book', *vol-i*

'eye', etc. would seem to originate form a syllabic requirement excluding final sequences *obstruent-sonorant*.[3]

(5) a. l-a bɔtʃ-a / l-a dʒamb-a / l-a maŋ
the-f mouth-f the-f leg-f the-f hand
'the mouth' 'the leg' 'the hand'

a'. l-i bɔtʃ-i-s / l-i dʒamb-i-s / li maŋ-s
the-pl mouth-pl-pl the-pl leg-pl-pl the-pl hand-pl
'the mouths' 'the legs' 'the hands'

b. il ko:r / fouk / il dint / il dʒal
the.msg heart / fire / the.msg tooth the.msg cock
'the heart /fire / the tooth the cock'

b'. i ko:r-s / fouk-s / i dintʃ / i dʒa-i
the.pl heart-pl / fire-pl / the.pl teeth.pl / the-pl cock-pl
'the hearts / fires / the teeth the cocks'

San Giorgio

If we look at the distribution of the plural inflections inside the DP, we see that the article is marked only by the *-i* ending in the feminine (6a'). Demonstratives may in turn lack *-s*, as in (6a'). The simple *i* plural inflection characterizes the masculine in the article and in the demonstratives, in (7a'), while other prenominal modifiers vary (*altr-i-s* vs. *bje-i*). Prenominal 1st/2nd/3rd person possessives set the plural feminine apart from the plural masculine by associating the feminine with *-s* and the masculine with *-i*, as in (6a') and (7a'). Finally (6a) and (7a) illustrate singular contexts and (6a") and (7a") illustrate the contexts with post-nominal adjectives.

(6) a. l-a / kist-a / kɛ bjel-a fɛmin-a l-a mɛ tʃamɛz-a
the-fsg / this-fsg / that fine-fsg woman-fsg the-fsg my.fsg shirt-fsg

a'. l-i / kist-i(-s) / kɛ bjel-i-s fɛmin-i-s l-i mɛ:-s tʃamɛz-i-s
the-pl / these-pl-pl / those fine-pl-pl women-pl-pl the-pl my.f-pl shirt-pl-pl

a" kɛ fɛmin-a grand-a kɛ(-s) fɛmin-i-s grand-i-s
that woman-fsg old-fsg those-pl women-pl-pl old-pl-pl

(7) a. il / kist-u / kel bjel tʃaŋ
the / this-msg / that nice dog
il ɲo tʃaŋ
the my.msg dog
'my dog'

a'. i / kist-i / ke-i bje-i tʃaŋ-s i ɲe-i tʃaŋ-s
the.pl / these-pl / those-pl nice-pl dog-pl the.pl my-pl tʃaŋ-pl
'the/these/those nice dogs' 'my dogs'

a". kɛl tʃaŋ gra:nt ke-i tʃaŋ-s graiŋ-s
that dog big those-pl dog-pl big-pl
'that big dog' 'those big dogs'

San Giorgio

[3] The status of *–i* in masculine singular lies outside the scope of this article; however, the point will be briefly picked up in Section 5.

The participles, in (8a,a'), for unaccusatives and transitives respectively, display the same inflectional system as nouns and adjectives, whereby feminine plural includes *-i-s* and masculine plural *-s*. In the paradigm of subject clitics, in (8b), *i* is associated to the 1st, 4th and 5th person. Object clitics, in (8a'), insert *i* in the plural forms, distinguishing masculine plural *i-u* from feminine plural *l-i*; however, the feminine plural form *l-i-s* is selected in final position of imperatives, as in (8c'). Moreover, *-i* characterizes also the dative clitic, which presents two alternants *i/g-i*, which can combine, doubling the lexicalization of the possessor, as in (8c'). In (8c') we label the *–l* morpheme generally occurring in definite determiners as Def.

(8) a. (lui) al e viɲu:t
he SCl.msg is come
'he has come'
(lo:r) a soŋ viɲu:-s / viɲud-i-s
they SCl are come-pl / come-pl-pl
'they have come'

(je) a ei viɲud-a
she SCl is.fsg come-fsg
'she has come'

a'. a l-u a kla'ma:t
SCl him-m have.3sg called.msg
'(s)he has called him'
a i-u a klama:-s
SCl them.pl-m have.3sg called-pl
'(s)he has called them (masculine)'

a l(-a) a klamad-a
SCl her-fsg have.3sg called.fsg
'(s)he has called her'
a l-i a klamad-i-s
SCl them-pl have.3sg called-pl-pl
'(s)hehas called them (feminine)'

b. i dur'mis 'SCl I.sleep, SCl you.sleep, etc.
ti dur'mis
al / a dur'mis
i dur'miŋ
i dur'mi:s
a dur'misiŋ

c. da- mi i-u / l-i-s
give to.me them.pl-m / them-pl-pl
'give me them'

c'. a i(-g-i) l-u da
SCl to.him/her/them it-m give.3sg
'(s)he gives him/ her it'
a i(-g-i) da kist-u-l
SCl to.her/him/them give.3sg this-m-Def

San Giorgio

2.1.2. Ladin (Rhaeto-Romance of Val Badia)

In Ladin varieties of Val Badia feminines and a subclass of masculines have the plural inflection *-s,* in (9a') and (9b'). Masculine determiners and a subset of the masculine nouns have *(-)i*, in (9b'). Plural masculine object clitics, in (9d), and dative, in (9d'), lexicalize as *i*. *-i* characterizes the Subject clitic of 3rd plural person, as in (9c). In the feminine plural in (9a) we tentatively identify the *-e-* morpheme inserted between the base and the inflection as a Class exponent.

(9) a. l-a dʒam-a
the-fsg
'the leg'

a'. l-e-s dʒam-e-s
the-Class-pl legh-Class-pl
'the legs'

b. əl kør / l ødl / l dʒa:l
the heart / the eye / the cock

b'. i kør-s / i ødl-i / i dʒa:-i
the.pl heart-pl / the.pl eye-pl / the.pl cock-pl
'the hearts / the eyes / the cocks'

c. a-i / al-əs do:rm
they.pl / they-Class-pl sleep
'they sleep'

d. i l / l-a / i / l-e-s la:f
SCl him / her-fsg / Dat / Def-Class-pl wash
'I wash him/ her/ them'

d'. da-i-l
give-Dat-Def
'give him it'

San Cassiano

2.1.3. Romansh

In Romansh varieties (Vattiz, Grisons), nouns, adjectives and determiners have the inflection *-s* in masculine and feminine plurals, cf. (10a, a') and (10b,b'); in participles, *-s* plural characterizes the feminine plural, in (10c), whereas the masculine plural has the *-i* inflection, as in (10d). As in the preceding systems, we assign gender and number properties to the inflectional morphemes on the basis of their distribution. So, for instance, the occurrence of *–a(-)* in singular and plural forms of feminines leads us to characterize it as F(eminine), independently of the number interpretation.

(10) a. l-a rɔd-a / l-a-s rod-a-s
the-f wheele-f / the-f-pl wheele-f-pl
'the wheel/the wheels'

a'. kwɛl-a-s bun-a-s dun-a-s
those- f-pl good- f-pl women-f-pl
'those good women'

b. iʎ kunˈti / iʎ-s kuntial-s
the knife / the-pl knife-pl
'the knife/the knives'

b'. kwel-s buŋ-s oməŋ-s
those-pl good-pl men-pl
'those good men'

c. ɛl-a ai veɲid-a / ɛl-a-s ai-n veɲid-a-s
she-f is come-f / they-f-pl are come-f-pl
'she has come / they have come'

d. el ai veɲi-u-s / el-s ain veɲi-i
he.m is come-f / they-pl are come- pl
'he has come / they have come'

Vattiz

2.2. Franco-Provençal

In the system of plural inflection in Piedmontese Franco-Provençal varieties, here investigated through the data from Cantoira (Val Grande di Lanzo), *-s* occurs in feminines, as in (11a'); masculine nouns lack any specialized morpheme for plural, as in (11b'). However *(-)i* characterizes masculine plural determiners and demonstratives, as in (11b'), the plural masculine OCl, in (11c), and the dative argument in (11d). The inflection of feminine singular is *–a* alternating with *-ə*, as in (11a), whereas masculine plurals are devoid of inflection, in (11b).

(11) a. l-a / sl-a fymɛll-ə vjəj-ə / l al-ə
the-f / that-f woman-f old-f / the wing-f
'the/that woman old / the wing'

a'. əl / səl fymɛll-ə-s vjəj-ə-s / əl-z al-ə-s
the-f.pl / that-f.pl women-f-pl old-f-p / the-f.pl wing-f-pl
'the/those women old/ / the wings'

b. l-u tʃa:t / sɔ-u tʃat lai
thre-m cat / that-m cat there
'the cat /that cat'

b'. i tʃa:t / sɛ-i tʃa:t lai
the.pl cat / those-pl cat there
'the cats / those cats'

c. l-u/lə/l-i/əl tʃammu
him-m/her-f/them-pl/them.f call-1sg

d. u i l-u də-nt
SCl to.him it-m give-3pl

Cantoira

2.3. Empirical generalizations

On the whole, a type of micro-variation emerges in which the presence of *-s* plural inflection does not exclude, but generally coexists with, the *(-)i* plural. (12) schematizes the distribution of the inflection *(-)i* for plural and dative in the varieties we have so far examined. Specifically, (12) compares *-i* and *-s* in nouns, N-i vs N-s, and registers the occurrence of *-i* in the other contexts. In the dialects, where *-i* or *-s* characterize only masculine or feminine plural, the specification m/f is introduced. All in all, we note that the specialized domain for *(-)i* are the Determiners and Clitics. Determiners are dedicated places for the specialized plural inflection, so that we can find dialects in which the masculine plural nouns are devoid of a specialized morphologization of plural, as in Cantoira, while Determiners lexicalize it as *-i*. Object clitics systematically externalize masculine plural and dative by means of *-i* inflection. The dialects endowed with subject clitics generally associate *(-)i* to the masculine plural. Participles present the distribution holding in nouns and adjectives except in the case of Romansh varieties, where *-i* inflection characterizes masculine plural only in participle. In (12) the subscript m/f specify the gender interpretation of the *(-)i* inflection.

(12) Plural *(-)i*

	D-i	N-i	(N-s)	Adj-i	SCl	OCl	DatCl	Participle-i
Vattiz	-	-	(+)	-				$+_m$
Cantoira	$+_m$	-	$(+_f)$	-	-	$+_m$	+	-
Comeglians	$+_m$	$+_m$	(+)	$+_m$	+	$+_m$	+	-
San Giorgio	+	+	(+)	+	+	+	+	$+_f$
San Cassiano	$+_m$	$+_m$	(+)		$+_m$	$+_m$	+	-
Ardauli	-	-	(+)	-			+	
Orroli	+	-	(+)	-			+	
Luras	-	-	(+)	-			+	

Another factor of variation has to do with the selection of a specialized Class inflection morpheme in the context of *-s*, as illustrated in (13). In the Friulian variety of San Giorgio *-i-* extends both to feminine and masculine, where it combines with *-s* marker, so generalizing the duplication of the plural inflection, as is the case in Sardinian for Orroli's *i-s*. It is of note that determiners separate the masculine plural, coinciding with the simple inflectional element *i* (cf. (5)-(8)), and the feminine plural form, where *-i* combines with the definiteness base *l-*. In North Sardinian Logudorese dialects, masculine plural selects a specialized Class Inflection vowel, i.e. *-o-*, as in Ardauli. As a consequence, the plural interpretation is doubled in two inflectional elements, *-s* and *-o-*. In the dialect of Luras. *-a-* lexicalizes plural, limitedly to determiners/demonstratives. In Cantoira's dialect in feminine plural the inflection *-ə-* occurs

(13) Specialized Class Inflection in sigmatic plural

	N		D
Vattiz			
Cantoira		+ $-e-_f$	
Comeglians			
San Giorgio	+ -i-		
S. Cassiano			
Ardauli		+ $-o-_m$	
Orroli			+ -i-
Luras	+ $-o-_m$		+ -a-

3. Morphological analysis

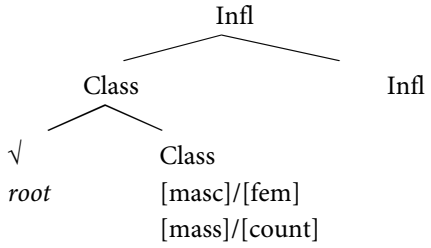
In what follows we will propose a model of the internal morphological organization of the noun based on the idea that inflectional elements are bona fide lexical entries endowed with interpretive content (Manzini and Savoia 2011b, 2017a,b, Savoia et al. 2017a,b). This theoretical assumption separates our approach to morphosyntax both from classical Distributed Morphology (DM) and from other models in which exponent are inserted under clusters of features subject to be manipulated by morphological rules. Specifically, as concerns the structure of the Noun, we assume that the first component is a root; following Marantz (1997), the root √ is category-less. Next to the root a vocalic morpheme encodes properties that, depending on the language, include gender and/or number and/or declension class. A third slot may be available, specialized for number (e.g. Spanish) or for case (e.g. Latin). In the literature (Picallo 2008, cf. Déchaine et al. 2014 on Bantu nominal classes, Fassi Fehri 2015 on Arabic) at

least two functional projections are needed in addition to the root – roughly gender and number.

In the approach we adopt, inflected nouns are analysed as the result of a Merge operation (Chomsky 2013) that combines a lexical root with Gender (feminine/masculine) and other classificatory properties including number, that contribute to specifying the argument introduced by the lexical root. Gender is selected by the root (cf. Acquaviva 2009, Kramer 2015) or freely combines with it determining a compositional interpretation.[4]The category-less lexical root is interpreted as a predicate, in the terms of Higginbotham (1985) and merges with the inflectional/derivational elements (Gender, number and in general classifiers), which may be thought of as predicates that restrict the properties associated to the argument *x* open at the predicate. In other words, these elements restrict the content of the argumental variable bound by Determiner/Quantifier.

The standard DM (Halle and Marantz 1993) treatment of inflectional class (Oltra-Massuet and Arregi 2005, Kramer 2015) has a Th(ematic vowel) node adjoined to Class/*n* post-syntactically. The content of Th are diacritics such as [I], [II], etc. for I, II inflectional class, etc. in turn spelled out as *-a, -o*, etc. (in Spanish, etc.). We do not adopt this treatment in so far as it presupposes post-syntactic Merge. We assign the inflectional morpheme to an Infl category, which syntactically merges with Class, including the root and its Gender specification. Class corresponds to Gender. The relation between inflectional elements, and Gender specifications can be understood in terms of selection by the inflection. Our various assumptions are schematized in (14).

(14)



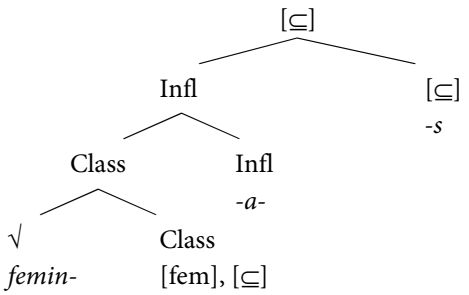
In Italian, the plural is obtained by a change of the inflection, differently from languages like Spanish and Sardinian, endowed with an independent lexicalization for the plural, namely *-s*. More precisely, we find alternations such as *ragazz-o* 'boy'/*ragazz-i* 'boys' in II class masculines, where the morpheme *-i* is specialized for plurality. Otherwise, the plural is externalized by the *-e* inflections, cf. *ragazz-a/ragazz-e* 'girl/ girls', and *-a*, cf. *dit-o/dit-a* 'finger/ fingers', independently occurring as Gender specification in the singular.

In Sardinian varieties, where *-s* of plural masculine combines with the specialized morphology *-o-*, contrasting with the *-u* ending of the singular, Class elements *-a*, *-o*, *-u* merge with the combination *root-Gender,* externalizing the inflectional class. The specialized morpheme *-s* in turn introduces the plural, as in (15). Following the proposal in Manzini and Savoia (2011b, 2017a,b) plural morphology is associated to the part-whole/inclusion content,

[4] In Romance languages, (root, Class) combinations interpreted compositionally are of the type *gatt-o* 'he-cat', *gatt-a* 'she-cat' in Italian. Some lass contents are determined directly by the root (see Italian *donn-a* 'woman', feminine or *marit-o* 'husband', masculine).

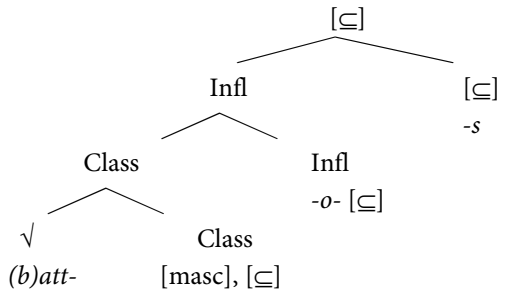
i.e. [⊆]. In other words, the content of the plural, [⊆], suggests that the argument of the root can be partitioned into subsets. In (15), [$_{\subseteq}$ s] merges with the set {{femin- [fem, ⊆]} *a*}, independently lexicalizing the singular, giving rise to the plural *femin-a-s*, which unlike the atomic singular has the property of being divisible (into subsets).

(15) cf. (1a'), (2a'), (3a')



The *-a* inflectional element in (15) or the *-u* inflectional element attribute specifications such as sexed interpretations (and other) to the argument required by the root. In the case of the plural masculine of *(b)att-o-s* 'cats' in (16), a specialized inflectional exponent is selected, i.e. *-o-*, evidently associated to the plural. As a consequence, the plural is doubled on two different inflections, a vocalic morpheme, usually connected to Class/inflectional properties, and *-s*, specialized for [⊆], as illustrated in (16). Tentatively we may conclude that *-a* and *-u* correspond to an individual interpretation in Sardinian, insofar as the plural is realized by *-s*. Since *-o-* emerges with *-s*, we may think that in the relevant varieties *-o-* is specialized for the plural interpretation.

(16) cf. (1b')



In Luras, the *-a-* inflection is inserted in plural determiners/adjectives, independently of Gender and Class inflection of the noun. We are led to conclude that *-a-* is available for the plural reading. In other words, *-a-* admits a plural interpretation. This recalls the fact that in many Romance varieties *-a* introduces plurals in a subset of nouns, externalizing a collective interpretation, as in (17a,b) for body parts in Standard Italian. In other dialects *-a* occurs in the singular as well as in the plural of feminine nouns, as in some North Italian dialects, illustrated in (18a) with the North Lombardy variety of Tresivio (Valtellina), and in North West Tuscany dialects (Garfagnana and Lunigiana), as in (18b).

(17) a. l-a cas-a l-e cas-e
the-fsg house-fsg the-fpl house-fpl
a'. il dit-o l-e dit-a
the.msg finger-msg the-fpl finger-f(pl)

(18) a. l-a femm-a i femm-a
the-f woman-f the.pl women-f
b. l-a femən-a
the-f woman-f (sg/pl)

Manzini and Savoia (2017a,b) argue for an analysis that assigns the [aggregate] interpretive property to the *-a* element, where [aggr] is understood as the conceptualization of a weakly differentiated set of individuals. The notion of aggregate is used by Chierchia (2010) to characterize the common core of mass and plural denotation. So, *-a* plural, instantiating the content [aggr], can be kept distinct from the *-i* plural. This raises the question how the [aggr] specification *-a* on N comes to agree with the [⊆] specifications that we have imputed to *-s*. One possibility is that [aggr], in that it introduces a type of plurality, is compatible with [⊆]; in effect the latter property is a subtype of the former. This analysis helps us treat the occurrence of *-a-* in the plural in *-s* as endowed with the interpretation [aggr], doubling the plural interpretation introduced by *-s*. This can be seen in the feminine plural in (19a) and in the masculine plural D in (19b).[5]

(19) a. [[[$_{\sqrt}$ femin] [$_{\text{Class}}$ fem]] [$_{\text{Infl [aggr]}}$ -a]] [$_{[\subseteq]}$ -s] cf. (3a')
b. [[[$_{\text{D}}$ s] [$_{\text{Class}}$ fem]] [$_{\text{Infl [aggr]}}$ -a]] [$_{[\subseteq]}$ -s] dið-ɔ-zɔ cf. (3b')

In Southern Sardinian, the plural element *i-* occurs in the definite article *i-s*. In these dialects, *-u* is the inflectional vowel of masculine nouns both in singular and plural. Differently from *-o-*, it is not specialized for the plural interpretation. On the contrary, in the article, we again find a double realization of the plural, combining *-s* with the *i-* plural element, as in (20).

(20) [[i $_{\text{Infl [⊆]}}$] s $_{[\subseteq]}$] cf. (2a',b')

An interesting point is that the doubling in (20) somehow preserves the double plural specification *-o-s* in the Northern dialects in (16) by different morphological means.

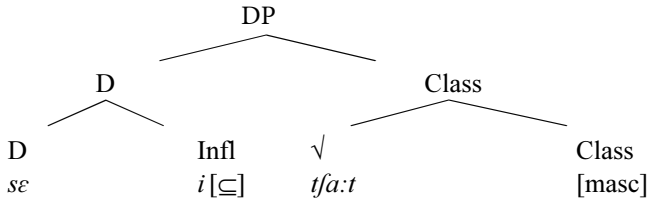
4. The interplay between *-s* and *-i*

In the Friulian, Ladin, Romansh and Franco-Provençal varieties in (5)-(11) we find a hybrid situation where the specialized *-s* inflection and *-i* coexist even while showing a partially different distribution. In Franco-Provençal (Cantoira), feminines preserve the *-s* plural. *(-)i* is restricted to the masculine plural of determiners, whereas masculine nouns lack inflection, as in (12). More precisely, masculine nouns lexicalize only the basic form both in the singular and in the plural like many North-Italian dialects (Rohlfs 1968 [1949], Manzini and Savoia 2005), as illustrated in (21). As a consequence, masculine nouns can be understood as including a class

[5] Manzini and Savoia (op.cit,) discuss the further issue whether *-a* is still [aggr] when appearing in the feminine singular. Needless to say one may stipulate that the [aggr] property is only optionally associated with *-a*. However, their proposal is that [aggr] is very much like [masc] or [fem]. Gender specifications are compositionally interpreted as referring to sex when taken together with animate roots – otherwise they will not be interpreted compositionally (they will be the equivalent of idioms). Similarly, [aggr] is compositionally interpreted with roots denoting mass referents or as plural.

property, agreeing with that of the determiners and modifiers, which, on the contrary, lexicalize [⊆] by means of the *-i* inflection.

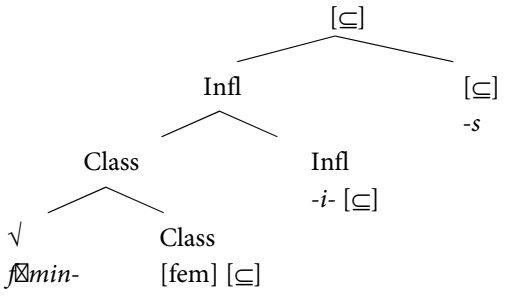
(21) cf. (11b)



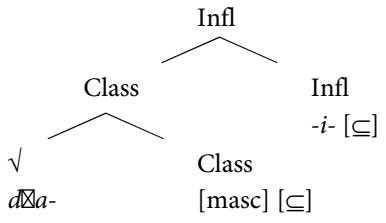
In Friulian (Comeglians, S. Giorgio) and Rhaeto-Romance (Vattiz, S. Cassiano), *-s* surfaces both in feminines and in masculines. The singular feminine has the inflection *-a*, the singular masculine is devoid of any specialized inflection. In Comeglians in (4) and in Vattiz in (10), where the feminine plural takes the form *-a-s*, *-a-* complies with the plural interpretation, in the terms discussed in relation to (19).

In Surselva Romansh (Vattiz), *-i* is confined to agreeing stative/unaccusative participles. In Friulian and Val Badia Ladin (S. Cassiano), *-i* or palatalized forms occur in the plural of a subset of masculines. Palatalized plurals imply an original final *-i*, as in [dintʃ] 'teeth', [bjei] 'fine.mpl', from something like *dint-i, bell-i*, etc. Furthermore *(-)i* occurs in nominal determiners, modifiers and in clitics. In the masculine, *(-)i* on the nouns and determiners is in complementary distribution with *-s*. In the San Giorgio variety, *-i-* is required also in feminines, so that the plural is reduplicated in feminine nouns, as in (22a). In the feminine, then plurality systematically requires *–s*; in the masculine, *-i-* is able to lexicalize plurality by itself. In short, in masculine plurals, only *-i* or only *-s* occurs, as in (22b) and (22c). In keeping with the analysis discussed above in (15), we associate *-s* with the specialized plural element [⊆]. As to *-i*, it is associated with the same content, although a slightly different denotation seems to be evoked by its being able to introduce also the possessor, as seen in (8 c,c'). We return to this issue at the end of section 5.

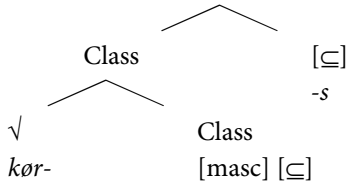
(22) a. cf. (6a')



b. cf. (5b')



c. cf. (9b')



The data highlight a general trend whereby feminine is basically endowed with richer denotational morphology than masculine, which, on the contrary, favours a more parsimonious gender and plural morphology. This holds for Friulian and in a general perspective (Rhaeto-Romance, Franco-Provençal), as summarized in (23).

(23)

	plural articles		plural object clitics	
	feminine	masculine	feminine	masculine
Comeglians	l-a-s	i	l-a-s	i-u
San Giorgio	l-i	i	l-i(-s)	i-u
San Cassiano	l-e-s	i	l-e-s	i
Cantoira	əl-s	i	əl-s	(l-)i

In the feminine, exponents with stronger referential import are generally preserved. Thus, in feminines, *-a*, which has a clear referential interpretation (sexed/ aggregate), systematically occurs. In the plural, feminines enhance this interpretive content by requiring a regular externalization of plurality by means of *-s* or possibly, of *–i*. Moreover, in many varieties, feminine articles and clitics require the *l-* lexical base, associated with definiteness, whereas this is not necessarily the case for masculines.

5. Asymmetries in plural specification and other restrictions

A property shared by the varieties reviewed here is that all or part of the plural specifications are realized on the determiners and possibly in prenominal modifiers but not necessarily on nouns. In the literature, the asymmetry between the agreement properties of determiners (and nominal modifiers/adjectives) vs nouns have been brought to light (cf. Cinque 2009). Indeed, different types of split emerge. By way of an example, Bonet et al. (2015) discuss cases in which prenominal determiners and adjectives lack (a set of) agreement properties. Their idea is that prenominal agreement is due to a "family of constraints" enforcing general agreement at PF; on the contrary, postnominal agreement is syntactic in nature and triggered by Spec-Head agreement.

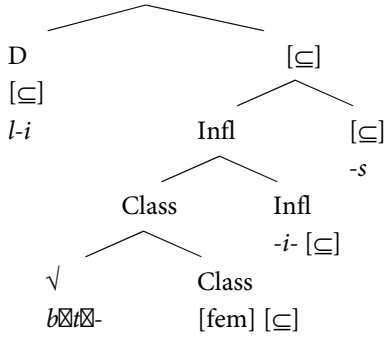
The hypothesis that different manifestations of agreement could be referred to different syntactic operations is pursued by several authors. Specifically, some approaches treat noun-modifier agreement (concord) as a process applying in the morphological component, separating it from the syntactic subject-verb agreement mechanism (Baier 2015). Nevins (2011: 8, 9) deals with the different behaviour of plural and gender agreement in DPs in several languages by assuming that the ability of number in ambiguously extending to object or subject descends from the underspecified treatment of singular, whereby "unmarked values of number, e.g. [-singular], are never syntactically active and never referred to in the syntax". By contrast,

"person features are always fully specified on syntactic arguments", thus excluding generalization processes. A mechanism based on the split between different types of features, namely marked vs. unmarked, is pursued in Pomino (2012) in accounting for the lack of number inflection in some Italian dialects.

Costa and Figueiredo (2002) present data concerning some Brazilian Portuguese varieties in which the plural inflection *-s* only occurs on the determiners o prenominal adjectives, as in *O-s/est-es/algun-s/un-s livr-o muit-o bonit-o* 'the/these/some book very nice'. Costa and Figueiredo adopt a distinction between dissociated and singleton morphemes. According to Embick and Noyer (2001), agreement and case morphemes are not syntactic projections and so they are not represented in syntax but they are added postsyntactically in morphology. Thus, dissociated morphemes convey an information "separated from the original locus of that information in the phrase marker" (Embick and Noyer 2001: 557). According to Costa and Figueiredo (2002), in European Portuguese, where plural agreement occurs on all elements inside DP, plural is a dissociate morpheme. The plural in Brazilian Portuguese corresponds to a specialized interpretable morpheme (singleton), which combines only with the "element anchoring the information concerning number", namely Determiners. Therefore, the Spec-head configuration, responsible for Subject-I agreement, is different from the D-N relation, where the plural singleton occurs.[6]

Our data call into question these approaches in that the asymmetries between determiners/ prenominal modifiers and nouns involve not so much the realization of plural inflection as the nature of the plural inflection. As we have proposed, the element *(-)i* can be understood as a specialized morpheme, associated to determiners, subject or object clitics and possibly prenominal modifiers. In present terms, the fact that *(-)i* systematically lexicalizes the plural independently of gender distinctions means that its only content is the plural property [⊆]. If we concentrate on the Friulian variety of San Giorgio in (5)-(9), we see that *-i* is able to lexicalize plural in any context, thus both on the determiner and the noun in (24) – though by itself only in determiners, *l-i* for feminines like (24) and *i* for masculines. Moreover, it is able to lexicalize plural in clitic occurrences, as exemplified in (8).

(24) cf. (5a')



[6] Following Vikner (1997), the presence of a specialized inflection in a sub-set of forms allows us to assume that the usual verbal agreement mechanisms apply. On this basis, Costa and Figueiredo conclude that despite the partial lack of Verb agreement in Brazilian Portuguese, it requires V-to-I movement and agreement just like European Portuguese.

In the minimalist framework (Chomsky 2001), agreement processes are associated with the rule of Agree – conceived for agreement in the sentential domain. We keep the assumption that Agree also applies within DPs. However, what impels Agree to apply is the necessity of creating equivalence classes of phi-feature bundles denoting a single referent (Manzini and Savoia 2007, 2011a,b, 2017a,b, Savoia et al. 2017). In this way, we avoid introducing the interpretable/uninterpretable distinction (or the valued/unvalued one), which are particularly difficult to implement within the DP, where both the lexical head N and its Determiners can stake good claims to interpretable/intrinsically valued status.

Our approach is based on a privative treatment of the morphological properties in lexical elements. As stated at the outset, it excludes postsyntactic manipulation or insertion of new material, keeping instead to the hypothesis that syntax is projected from morphological terminals which have interpretive lexical content. As a result, the distinction between morphological agreement (dissociated) and syntactic agreement has to be rephrased as the split between agreement inflections with general occurrence and those with restricted distribution. As already mentioned, the distribution of *-s* and *(-)i* occurring complementarily or in combination, excludes the notion of dissociated morpheme as a morpheme generalized by externalization. Thus, in systems like those of San Cassiano in (8), San Giorgio in (7) and Comeglians in (4), where *-s* occurs also in the masculine, the plural masculine determiners have the *-i* inflection, co-occurring however with the ending *-s* in the noun, as in *i kørs* 'the hearts' (San Giorgio).

In other words, we witness a type of morpho-syntactic split, whereby definiteness and deictic elements are endowed with specialized morphology, imputable to the role they play in the identification of arguments. In general, the occurrence of specialized plural elements is associated with the head of DP phase, i.e. determiners D and possibly other nominal modifiers, as schematized in (25a). In the vP phase, schematized in (25b), object clitics generally lexicalize plural specifications by means of the specialized exponent *(-)i*. This fits in with the idea that object clitics are the true agreement head of the vP phase (Roberts 2010, Manzini et al. forthcoming). The selection of *-i* in Romansh participles in (10) can be in turn explained as an instantiation of Object-v agreement, in varieties lacking object clitics, as in the case of Surselva Romansh. Finally, subject clitics in many varieties do not register a specialized plural reading. It is natural to relate this to the presence of a pronominal inflection on the finite verb in I heading the CP-phase in (25c)[7].

(25) Friulian / Ladin / Romansh varieties: masculine plural

a.	DP phase:	D (Q)	Adj	N
		-i	-s/-i	-s/-i
b.	vP phase:	ObjCl	v (Participle)	
		(-)i	-s/∅/-i(Romansh)	
c.	CP phase:	SubjCl	I	
		∅/(-)i	verbal Infl	

[7] Manzini et al. forthcoming propose to connect the distribution of the plural inflections with a parameter whereby an agreement property can be differently externalized on the head of the phase vs the complement of the phase.

A question we have left out so far is the doubling of the plural by *-i-* and *-s*. This suggests that in Romance *-s* and *-i* have a slightly different content. This conclusion is upheld by the fact that *-i* is also the dative, so introducing an interpretation [⊆] including also the possessive relation between possessor and possessed. Manzini and Savoia (2011a,b) connect dative to the inclusion, expressed by [⊆], in which the dative is the locator or the possessor of the internal argument of the verb. In other words, the Romance dative *(-)i* inflection is able to take scope both over the nominal root, giving rise to the plural interpretation, and over the indirect object – internal argument pair, giving rise to the dative reading. On the contrary, *-s* introduces the part-whole relation only on the nominal root (the plural); in this sense it. contrasts with the less specialized interpretation of *-i*, that in fact extends to the possessive inclusion reading. We see that in the instances of doubling, it is the more specialized, hence richer, content of *-s* that takes scope over the less specialized, hence simpler *-i* (cf. (19a)). Finally, the quantificational properties of *-i* are compatible with singular in a subset of masculine nouns in Friulian (cf. San Giorgio in (5)), cf. fn. 3.

6. Conclusions

This work deals with the question of the theoretical status of nominal inflections, specifically plural inflections, in some Romance varieties. Our intent is to better understand the interface between the computational system and externalization, the traditional syntax/morphology interface. The microvariation in the plural inflection that we have examined brings into light the interplay between *-i* and *-s* as plural exponents. Some general points arise as regards the distribution of *-i*:

- *(-)i* plural typically appears on determiners and modifiers of the noun and in a subset of plural masculine nouns.
- *(-)i* plural is lexicalized in subject and object clitics
- *(-)i* lexicalizes dative
- *(-)i* may be associate also to feminine nouns (San Giorgio)
- *(-)i* can combine with *-s* giving rise to doubling of plural morphology: this suggests a difference in the content of the two inflections.

We have connected these facts with the different nature of *-i* and *-s* in the light of a theoretical model where inflectional morphemes are endowed with interpretable content, on the basis of which syntactic structures and computations are built. In this perspective, agreement is the result of the equivalence between phi-feature bundles denoting a single referent. Our approach allowed us to account for the distribution of *-i* by tying it to the requirements of the determiners and in general the modifiers of nouns. More precisely, we have connected *-i* with the definiteness and deictic elements corresponding to the identification of arguments and playing the role of agreement head of phase. The association of *-i* with dative suggests a partially different content from *-s*, so explaining their doubling

References

Acquaviva, Paolo. 2009. "Roots and Lexicality in Distributed Morphology". In *York Working Papers – Special Issue York-Essex Morphology Meeting May 2009*, edited by Alexandra Galani, Daniel Redinger and Norman Yeo, 1-21.

Baier, Nico. 2015. "Adjective Agreement in Noon: Evidence for a Split Theory of Noun-Modifier Concord", *Paper presented at LSA 2015, Portland*.

Béjar, Susana and Milan Rezac. 2009. "Cyclic Agree". *Linguistic Inquiry* 40: 35–73.

Blasco Ferrer, Eduardo and Michel Contini. 1988. "Sardisch: Interne Sprachgeschichte I. Grammatik". In *Lexikon der Romanistischen Linguistik*. IV: *Italienisch, Korsisch, Sardisch.* edited by Günter Holtus, Michael Metzeltin and Christian Schmitt, 836-853. Tübingen: De Gruyter.

Blevins, James. 2006. "Word-based morphology". *Journal of Linguistics* 42. 531-573.

Bonet, Eulàlia and Joan Mascaró, 2012. "Asimetrías de concordancia en el SD: el rasgo de masa en asturiano". In: Antonio Fábregas, Elena Felíu, Josefa Martín & José Pazó (eds.), *Los límites de la morfología. Estudios ofrecidos a Soledad Varela Ortega*, 91-104. Universidad Autónoma de Madrid: Servicio de Publicaciones de la Universidad Autónoma de Madrid.

Chierchia, Gennaro. 2010. "Mass nouns, vagueness and semantic variation". *Synthese* 174. 99–149.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2001. "Derivation by Phase". In *Ken Hale: A Life in Language,* edited by Michael Kenstowicz, 1–54. Cambridge, MA: The MIT Press.

Chomsky, N. 2005. "Three Factors in Language Design". *Linguistic Inquiry*, 36: 1–22

Chomsky, Noam. 2013. "Problems of projection". *Lingua* 130. 33–49.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174. 99–149.

Cinque, Guglielmo. 2009. "The fundamental left-right asymmetry of natural languages". In Sergio Scalise, Elisabetta Magni and Antonietta Bisetto (eds.), *Universals of language today*, pp. 165-184. Dordrecht: Springer.

Costa, João and M. Cristina Figueiredo Silva. 2002. "Notes on nominal and verbal agreement in Portuguese". *Rivista di Grammatica Generativa*, 27: 17-29.

Déchaine, Rose-Marie, Raphaël Girard, Calisto Mudzingwa and Martina Wiltschko. 2014. "The internal syntax of Shona class prefixes". *Language Sciences* 43. 18-46.

Embick, David and Rolf Noyer. 2001. "Movement Operations after Syntax". *Linguistic Inquiry* 32: 555-595.

Fassi Fehri, Abdelkader. 2016. "Semantic gender diversity and its architecture in the grammar of Arabic". *Brill's Journal of Afroasiatic Languages and Linguistics* 8: 154-199.

Halle, Morris and Alec Marantz. 1993. "Distributed morphology and the pieces of inflection". In *The view from Building 20,* edited by Kenneth Hale and Samuel J. Keyser, 111-176. Cambridge, MA.: The MIT Press.

Higginbotham, James. 1985. "On Semantics". *Linguistic Inquiry* 16. 547-621.

Jones, Michael. 1993. *Sardinian syntax*. London: Routledge.

Kramer, Ruth. 2015. *The Morphosyntax of gender*. Oxford: Oxford University Press.

Lausberg, Heinrich. 1971. *Linguistica romanza. II Morfologia*. Milano: Feltrinelli.

Manzini, M. Rita and Leonardo M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*, Alessandria: Edizioni dell'Orso,

Manzini, M. Rita and Leonardo M. Savoia. 2007. A Unification of Morphology and Syntax. Investigations into Romance and Albanian dialects. London: Routledge.

Manzini, M. Rita and Leonardo M. Savoia. 2011a. *Grammatical Categories*. Cambridge: Cambridge University Press.

Manzini, M. Rita and Leonardo M. Savoia. 2011b. Reducing 'case' to denotational primitives: Nominal inflections in Albanian. in Linguistic Variation 11, 76-120.

Manzini, M. Rita and Leonardo M. Savoia. 2017a. "N morphology and its interpretation: The neuter in Italian and Albanian varieties". In Anna Bloch-Rozmej and Anna Bondaruk (eds.), *Constraints on Structure and Derivation in Syntax, Phonology and Morphology*, pp. 213-236. Frankfurt: Peter Lang.

Manzini, M. Rita and Leonardo M. Savoia. 2017b. "Gender, number and inflectional class in Romance". In Joseph Emonds and Markéta Janebova (eds.). *Language Use and Linguistic Structure*, pp. 263-282, Olomouc: Palacky University Olomouc.

Manzini M. Rita and Leonardo M. Savoia, forthcoming. "N morphology and its interpretation: Romance feminine/plural –a". In Silvio Cruschina (ed.), *Proceedings of CIDSM 2016*.

Manzini M. Rita, Benedetta Baldi and Leonardo M. Savoia, forthcoming. "Asymmetries in the positioning of plural morphology in DPs in *-s* and *-i* systems: A Romance comparative perspective". In *Proceedings of CIDSM* 2018.

Nevins, Andrew. 2011. "Multiple agree with clitics: person complementarity vs. omnivorous number". *Natural Language and Linguistic Theory*, 29: 939-971.

Marantz, Alec. 1997. "No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon". *University of Pennsylvania Working Papers in Linguistics* 4. 201–25.

Meyer-Lübke W. 1899. *Grammatik der Romanischen Sprachen. Dritter Band: Syntax*. Leipzig

Molinu, Lucia. 2013. "Riflessioni sulle realizzazioni di /-s/ in alcune parlate sarde confrontate con quelle di alter varietà romanze", contribution presented at XXVIIe Congrès International de Linguistique et de Philologie Romanes, Nancy, 15 –20 Juillet 2013.

Oltra-Massuet, Isabel and Karlos Arregi, 2005. "Stress-by-Structure in Spanish". *Linguistic Inquiry* 36(1). 43-84.

Picallo, Carme. 2008. "Gender and number in Romance". *Lingue e Linguaggio* VII. 47-66.

Pomino, Natascha. 2012, "Partial or complete lack of plural agreement: The role of morphology". In Sacha Gaglia and Marc Olivier Hinzelin (eds.), *Inflection and Word Formation in Romance Languages*, pp. 201-230, Amsterdam: John Benjamins.

Roberts, Ian. 2010. *Agreement and Head Movement*. Cambridge, Mas.: The MIT Press.

Rohlfs, Gerhard. 1968 [1949]. Grammatica storica della lingua italiana e dei suoi dialetti. Morfologia. Torino: Einaudi.

Sanna, Antonio. 1975. *Il dialetto di Sassari (e altri saggi).* Cagliari, Edizioni 3T.

Sauzet, Patrick. 2012. "Occitan plurals. A case for a morpheme-based morphology". In Sacha Gaglia and Marc Olivier Hinzelin (eds.), *Inflection and Word Formation in Romance Languages*, pp. 179-200, Amsterdam: John Benjamins.

Savoia, Leonardo M. 2015. I dialetti italiani. Sistemi e processi fonologici nelle varietà di area italiana e romancia. Pisa, Pacini.

Savoia, Leonardo M., M. Rita Manzini, Ludovico Franco and Benedetta Baldi. 2017a. "Nominal Evaluative Suffixes in Italian". In Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino and Anna Maria Thornton (eds.), *Di tutti i colori. Studi linguistici per Maria Grossmann*, pp. 283-300. Utrecht: Utrecht University Press.

Savoia, Leonardo M., M. Rita Manzini, Benedetta Baldi and Ludovico Franco. 2017b. "A Morpho-syntactic Analysis of Evaluatives in Italian". In *Studi Italiani di Linguistica Teorica e Applicata*, vol. XLVI. 413-440.

Vikner, Sten. 1997. "V°-to-I° movement and inflection for person in all tenses". In *The New Comparative Syntax*, edited by Liliane Haegeman, 189-213. Edinbourgh: Longman Linguistics Library.

LingBaW *Linguistics Beyond And Within 4 (2018), 161-172*

The interaction between relevant-set based operators and a topic–predicate dimension*

Anna Szetelia and Gábor Alberti
University of Pécs, Hungary

Abstract

Hungarian relevant-set based operators, such as universally quantified noun-phrases and the *also*-quantifier, signal a logico-pragmatic relation between their explicit meaning and a broader implicit set of relevant participants which property can be mentioned as "double referentiality" of the operator. Furthermore, they indicate the new or correcting information in a topic–predicate dimension which belongs to the broader world of the discourse. Our research aims to identify the differences by investigating the suprasegmental features of *each*-quantifiers and *also*-quantifiers on the Hungarian left-periphery indicating the two functions in the topic–predicate dimension. Short sentences in which quantifiers functioning either as the main (new/correcting) information or as a topic (anchor of relevant information) were read by 41 non-linguist native Hungarian subjects. The pitch, intensity and duration properties were analyzed by Praat. On the basis of paired t-tests of the data, it can be stated that a two-dimension model of information structure is required to capture the phenomenon.

Keywords: information structure, quantifiers, prosody, Praat, information focus

1. Introduction

Since the leitmotif of *Linguistics Beyond and Within* 2017 was *The Outskirts of the Regular*, this paper discusses noun phrases modified by the universal determiner *mind(en)* 'every', called universally quantified noun phrases by É. Kiss (2002). In contrast to general noun phrases (1a–a'), they do not trigger two types of *verbal prefix – verb stem* order (1b-b').

(1) a. Lilla el-jött.
Lilla VM-come.Past.3Sg
'Lilla has come.'
a'. Lilla jött el.
Lilla come.Past.3Sg VM
'It was Lilla who has come.'

* The present scientific contribution is dedicated to the 650th anniversary of the foundation of the University of Pécs, Hungary. We are grateful for the financial support of the University of Pécs (EFOP 343).

b. Mindhárom barátom el-jött.
All_three friend.Poss.1Sg VM-come.Past.3Sg
'All three of my friends have come.'

b' * Mindhárom barátom jött el.
All_three friend.Poss.1Sg come.Past.3Sg VM

Expressions modified with *csak* 'only' or *nemcsak* 'not only' (2a) seem to behave similarly in the sense that they tolerate only the *verbal prefix – verb stem* word order.

(2) a. * Csak / Nemcsak Lilla el-jött.
Only / Not_only Lilla VM-come.Past.3Sg

b. Csak / Nemcsak Lilla jött el.
Only / Not_only Lilla come.Past.3Sg VM
'Only / Not only Lilla has come.'

Nevertheless, the quantifier, which never evokes a *verb stem – verbal prefix* order, can have a topic (3) or a predicate/comment function (4) in a prosodic and pragmasemantic point of view.

(3) a. Köszönöm érdeklődő kérdésedet! Úgy jellemezném a barátaimat, hogy Anti imádja a teniszt, Béci szeret sakkozni és pingpongozni,
'Thank you for your inquiring question! I can characterize my friends the way that Anti likes tennis, Béci gladly plays chess and table tennis,'

b. *Csaba is* szeret pingpongozni, Top$_{\exists\checkmark}$
Csaba also like.3Sg tabletennis.Inf,
'Csaba also gladly plays table tennis,'

c. Anti és Csaba lelkes komolyzene-rajongó, Béci és Csaba gyakran söröznek,
'Anti and Csaba are great fans of classical music, Béci and Csaba often drink beer'

d. és *mindhárom* barátom oda-van Scarlett Johanssonért. Top$_{\forall\checkmark}$
and all_three friend.Poss1Sg VM-be.3Sg Scarlett Johansson.Cau
'and *all three of* my friends are crazy about Scarlett Johansson.'

(4) a. Azt gondolod, hogy csak Béci szeret pingpongozni, és csak Anti van oda Scarlett Johanssonért?
'Do you think that it is only *Béci* who gladly plays table tennis, and it is only *Anti* who is crazy about Scarlett Johansson?

b. *CSAba is* szeret pingpongozni, Pred$_{\exists\checkmark}$

c. és *MINDhárom* barátom odavan Scarlett Johanssonért. Pred$_{\forall\checkmark}$

As for the structure of the paper, the following section elaborates on the theoretical background of quantifiers and information focus and provides an appropriate model to capture the dual function of the relevant-set based operators. Then comes a section in which the statistically significant differences in prosodic properties of the quantifier types mentioned above are described, on the basis of Alberti and Szeteli's (2017) methodology. They are differentiated along the lines of old/given and new/predicative information (Szűcs, 2016), based on an experiment with the participation of 41 non-linguist native Hungarian subjects. The paper concludes with a section which, in addition to some concluding remarks, presents a further phenomenon in favor of our two-dimension operator model.

2. Towards a two-dimension model of Hungarian information structure

This part is devoted to the question as to which pre-V part of the Hungarian sentence *each*-quantifiers and *also*-quantifiers belong to.

The current stances are (still) all essentially based on É. Kiss (2002) (see also Szabolcsi, 1997; Alberti and Medve, 2000). According to Laczkó (2016, p. 217), both types of (non-contrastive) quantifier belong to the *predicate portion* of the sentence (Table 1), while Varga (2016, p. 46) places them in the zone that he calls the *Comment* (Table 2).

Table 1: *Hungarian sentence articulation, according to Laczkó (2017, p. 60)*

TOPIC	**PREDICATE**				
(A) (contrastive) topic, sentence adverb	(B) quantifier	(C) focus/VM		(D) verb	(E) postverbal constituents
		(Ca) focus	(Cb) VM		

Table 2: *Structural positions in the Hungarian Sentence, according to Varga (2016, p. 46)*

Sentence	
	Comment
T*	Dist* { (F) / (PDMA) (VMod) } V PostV*

As can be seen in the two tables, the two zones mentioned above, preceded by the topic zone (cf. T*), are practically the same, since the formula Dist* in Varga's (2016, p. 46) approach refers to nothing else but a sequence of quantifiers. Note in passing that the reason for this abbreviation is that it is a prominent property of quantifiers that they are to be interpreted in a *distributive* manner. In the sentence *mindkét fiú felemelte a zongorát* 'both boys lifted the piano,' for instance, the natural group-interpretation according to which the two boys lifted the piano together as a group is excluded. Only the much more artificial (distributive) reading according to which one boy lifted the piano alone and the other also lifted it alone is available.

An *each*-quantifier, very rarely, is analyzed in the literature as a topic, but only as a contrastive topic (Laczkó, 2016, p. 216, Varga, 2016, p. 74).

At this point it is worth making some words on contrastive topics. In Hungarian, the scope order of preverbal (non-*in-situ*) constituents corresponds to their surface order (5a-a'); as shown by the proposed paraphrase of (5b), however, quantifiers in (Spec,CTopP), marked by a special intonation, which the symbol '^' refers to in (5b), give the impression of having inverse scope, apparently violating this generalization (É. Kiss, 2002, p. 25).

(5) Illustration of the inverse-scope puzzle in Hungarian

a. **KE**vés diák olvasott <el> MINden regényt ?<el>.
few student read.Past.3Sg away(VM) each novel.Acc away(VM)
'It holds only for few students that the given student has read each novel.'

a'. MINden regényt **KE**vés diák olvasott el.
each novel.Acc few student read.Past.3Sg away(VM)
'It holds for each novel that few students have read it.'

b. ^**MIN**den regényt # **KE**vés diák olvasott el.
each novel.Acc few student read.Past.3Sg away
Meaning: practically the same as that given in (5a), and not that given in (5a')

This paper is not concerned with such type of quantifier; the interested reader is referred to Farkas and Alberti (2017). Note that such an approach is a preferable alternative, for those accepting multiple spell-out (e.g. Uriagereka, 1999) first applied to Hungarian by Surányi (2009). Now we intend to verify that both *each*-quantifiers and *also*-quantifiers can also serve as non-contrastive topics. Hence, our proposal is based on a two-dimension operator model.

One dimension is a relevant-set based (logico-pragmatic) tier (see Table 3). Roles in this tier can be identified on the basis of the system of the five types of logico-pragmatic content given in the table as follows. If the reference of a noun phrase is associated with a particular operator character in an utterance, then, by referring to r, a whole set of its pragmatic alternatives is evoked as background knowledge shared by the interlocutors. Such alternatives are thus not referred to explicitly, but only implicitly. Due to the given operator, some logical claim is predicated of the implicit referents.

Table 3: *The system of operators in Hungarian*

	✓	¬
∃	Q∃: *also*-quantifier	CTop: contrastive topic
∀	Q∀: *each*-quantifier	Foc: (contrastive) focus
	TopP: (non-contrastive) topic	

In all five examples shown in Table 4, the set of implicit referents consists of persons who can be regarded in a given context as alternatives to a person who is called Lilla. They all together form the *relevant set*. Suppose the implicit participants are Anna, Bea and Cili; so the relevant set now consists of four people.

Table 4: *Illustration of the system of operators in Hungarian*

	✓	¬
∃	LILla is EL- jött. Lilla also away(VM) came 'Lilla also came here.'	**^LIL**la # **EL**- jött. Lilla away(VM) came 'As for Lilla, she came here [but there is another person who did not come here].'
∀	MINdenki EL- jött. everyone away(VM) came 'Everyone came here.'	**LIL**la jött el. Lilla came away(VM) 'It was Lilla who came here.'
	LILla EL- jött. Lilla away(VM) came 'Lilla came here.'	

The corresponding sentence with an *also*-quantifier then provides the additional semantic information—in addition to the "explicit content" that Lilla came here, which is true in all the five variants—that what holds for Lilla also holds for (✓) at least one (∃) implicit participant. The additional information due to the contrastive topic is that what holds for Lilla does not hold for (¬) at least one (∃) implicit participant.

In certain contexts, there is (said to be) a "weaker" interpretation (eg., Gyuris 2009), which can be regarded as the application of the above-sketched strict logical interpretation but to an epistemic variant of the given proposition:

(5) ^**LIL**la **EL**- jött.
Lilla away(VM) came
'As for Lilla, *I am sure that* she came here [but there are other persons in whose case *I am not sure that* they came here].'

The contribution of (contrastive) focus is captured in Table 4 as follows: what holds for Lilla is a piece of information that uniformly (∀) does not hold for (¬) the implicit participants.

The *each*-quantifier realizes the fourth logical possibility in the following sense: everyone is referred to implicitly (since the general expression *mindenki* "everyone" can have no other function in the given context than evoking what is termed above the relevant set), and hence the corresponding sentence can be interpreted as claiming that the information "someone came here" holds truly (✓) for each implicit participant (∀).

As for the fifth operator, the non-contrastive topic, it can be placed in the system just sketched as an operator realizing the logical alternative of providing no information on the implicit participants. The translations illustrate these (context-based) semantic contributions. As for the formal cues of these operators, relative to the basic variant with a topic, the two types of quantifier can be recognized by means of characteristic elements such as *is* 'also' and the prefix *mind-* 'each'. The two contrastive operators can be recognized relying on peculiar intonational and word-order phenomena. The contrastively topicalized element bears a special rising and then falling intonation (^) and is followed by a short pause (#). The focused element bears a strong **FOCUS STRESS** and seems to substitute for the preverb compared to the neutral word order.

The other dimension is the topic–predicate, or topic–comment, tier, to be determined in the broader world of the discourse in a basically Büringian (1997; 2003) style (cf. Gécseg, 2013; Szűcs, 2017) considering the tradition of Bródy (1991).

The next table presents the connection between the two dimensions.

Table 5: *Relevant-set based pragmatico-logical contributions and topic–predicate status*

	✓			¬
∃	*is* 'also'	*is* 'also'	*nemcsak* 'not only'	Contrastive topic
∀	*mind* 'each'		*mind* 'each'	Narrow focus (identification by exclusion)
∅	Topic			Verbal modifier

The coloring intends to demonstrate that the dimension of relevant-set based/ "double" referentiality and the dimension of the topic–comment tier is only partially independent. Only the *each*- and *also*-expressions produce real independence: they can serve as both topics, serving as anchors for new information, and predicates/comments, providing this new information. *Nemcsak*-expressions and identificational foci (whose contribution is often characterized as "identification by exclusion" (É. Kiss, 1998)), however, can appear only as predicative elements left-adjacent to the verb stem performed with a deleting stress. Of the elements showing no double referentiality, normal topics are only of a topic character while verbal prefixes/modifiers triggering no deleting stress can function only as part of the comment portion of the sentence. As for contrastive topics, they do not occur in the predicative/comment zone in Hungarian, in spite of the fact that their Janus-faced logical character suggests a potential occurrence in the predicate/comment zone. A contrastive topic can appear in different pre-V parts of the sentence, but there are no arguments against a hypothesis according to which these potential places are all inside the topic portion. The occupation of the two extremities of the topic portion by a contrastive topic is illustrated in (6a-b):

(6) a. /PEtit bezzeg a drágalátos haverod \LIlinek mutatta be.
Peti.Acc [CTop] the precious friend.Poss.2Sg Lili-Dat introduce.Past.3Sg in(VM)
'As for Peti, your precious friend has introduced him to Lili.'

b. A drágalátos haverod persze /PEtit \LIlinek mutatta be.
the precious friend.Poss.2Sg of_course Peti.Acc Lili-Dat introduce.Past in(VM)
'As for Peti, your precious friend has introduced him to Lili.'

3. Suprasegmental features of the Hungarian relevant-set based operators

In the experiments, we attempted to transfer the appropriate reading of the *each*-quantifier and the *also*-quantifier into our subjects' minds. Therefore, the whole text presented in (3-4) was offered as a verbal stimulus for them. Then the relevant sequences were cut out and analyzed by Praat.

3.1. Methods of the experiment

The 41 non-linguist native Hungarian subjects, all first-year students of the University of Pécs, took part in the research voluntarily. They were not paid for their participation in the experiment. We have not intended to regulate the ratio of females to males, which was 35:6; it is postponed to future research to attempt to identify gender-dependent features in the relevant area (NB: a conjecture is that what males tend to express by means of stronger intensity is rather expressed by means of higher pitch by females).

After studying the short dialogue parts in (3-4) laid out in textboxes through 1 or 2 minutes, the participants told the conductor of the experiment that they have already been prepared for reading out the given texts. Since the goal of the research was to imitate spontaneous speech as truly as possible, the following instructions were given.

1. Please read out, or more precisely, play the following discourse parts by using a tone which is not monotonous; project yourselves to the situation on the basis of the given linguistic context.
2. Read out every single word in the word order given, because even tiny details may be crucial.
3. While playing the situation, consider that the punctuation is not completely given. Only dots and question marks have been used (NB: we must not have used the elements of punctuation immediately referring to tonal features, e.g., exclamation marks). Thus the participants had to read out the parts as they expected themselves to do that in an everyday communication situation.

The relevant utterances were embedded in a broader linguistic research in which different verbal stimuli were used to trigger different interpretations in the subjects' minds, following the methodology applied by Alberti and Szeteli (2017) to differentiate functions of discourse markers with the same phonetic forms constituting polysemous networks.

The experiments were recorded by a dictaphone (44,1 kHz/16 bit). The relevant fragments were analyzed by Praat.

3.2. Distinguishing the two information-structure based functions

The suprasegmental properties of the fragments presented in (3-4), repeated here as (7-8) were investigated along three parameters. Duration values of the single words, intensity values of their first syllables, and pitch values of the vowels of the first syllables were measured.

(7) a. *Csaba is* szeret pingpongozni, Top∃✓
Csaba also like.3Sg tabletennis.Inf,
'Csaba also gladly plays table tennis,'

b. és *mindhárom* barátom oda-van Scarlett Johanssonért. Top∀✓
and all_three friend.Poss1Sg VM-be.3Sg Scarlett Johansson.Cau
'and *all three of* my friends are crazy about Scarlett Johansson.'

(8) a. *CSAba is* szeret pingpongozni, Pred∃✓

b. és *MINDhárom* barátom odavan Scarlett Johanssonért. Pred∀✓

From the duration data presented in (9), it can be claimed that the elements with a predicative character, compared to the other words of the fragment, proved to be longer than the homophonous elements with a topic character. These elements considered are the universal-quantifier determiner *mindhárom* 'all three' in (8b) and the proper noun *Csaba* in (8a), which is the word bearing the main stress instead of the particle *is* 'also'.

(9) Mean of the duration values in milliseconds
a. The element of a Top∃✓-character: 239 ms, with a 1167 ms long continuation of the fragment
a'. The element of a Pred∃✓ character: 250 ms, with a 1039 ms long continuation
b. The element of aTop∀✓ character: 502 ms, with a 778 ms long continuation
b'. The element of a Pred∀✓ character: 522 ms, with a 738 ms long continuation

Figures 1-4 illustrate that different tendencies between the two readings could be measured in the intensity values as well as in the pitch values in the whole set of the subjects.

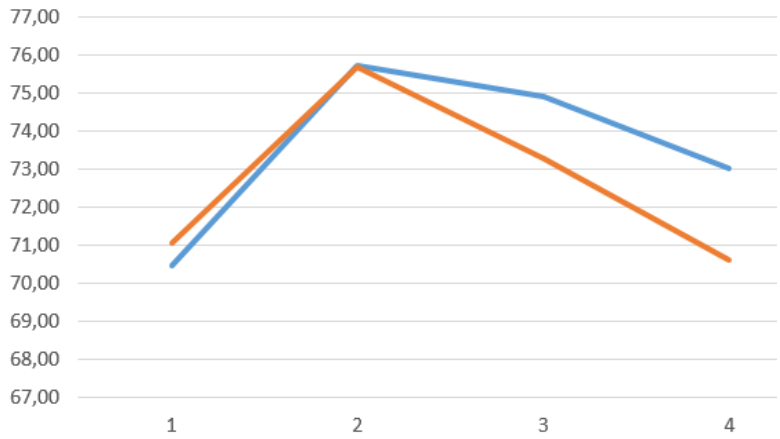


Figure 1: *Mean values of the intensity (in dB) in the case of the each-quantifier in topic context (blue) and in predicative context (orange) (n=41)*

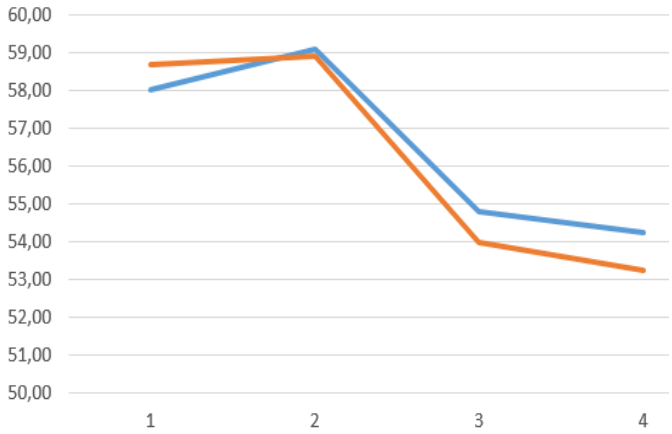


Figure 2: *Mean values of the pitch (in semitones, see Matteson et al. 2013) in the case of the each-quantifier in topic context (blue) and in predicative context (orange) (n=41)*

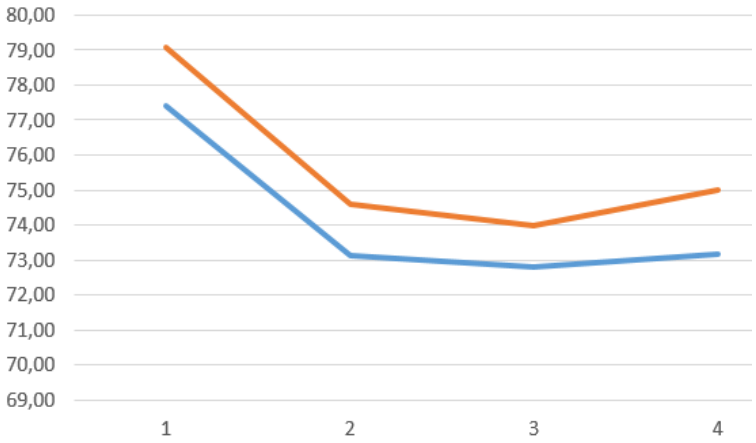


Figure 3: *Mean values of the intensity (in dB) in the case of the also-quantifier in topic context (blue) and in predicative context (orange) (n=41)*

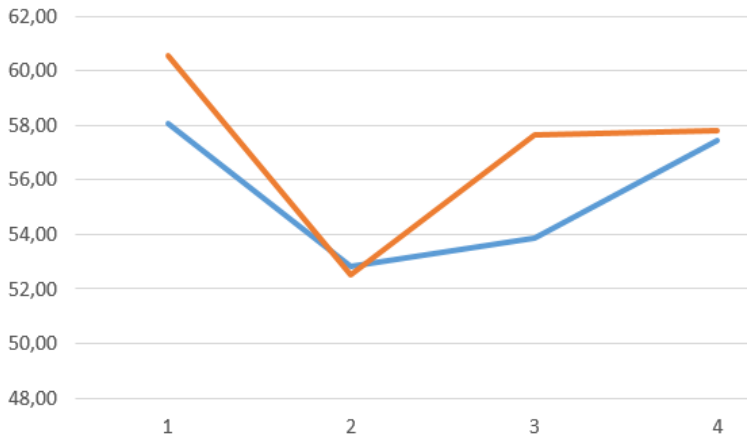


Figure 4: *Mean values of the pitch (in semitones) in the case of the also-quantifier in topic context (blue) and in predicative context (orange) (n=41)*

In addition to the whole set of experimental subjects, a subgroup of 31 was also investigated. Participants have been excluded for two reasons: (i) they did not pronounce all elements of the fragments to be read out, and/or (ii) their irregular voicing made it difficult to identify the pitch values (see McGlone, 1967; Hollien and Wendahl, 1968; Blomgren et al. 1998 cited by Markó, 2013, p. 19).

F-tests and paired t-tests were applied to prove that the differences between the multitudes are significant. Our goal was to differentiate the alternative readings of the two quantifiers, more or less in the same way. Thus we can formulate tendencies concerning the location of operators in the topic–predicate dimension.

In Table 6, the duration data were considered along with the ratio $m=l_q/(l_f-l_q)$ mentioned above, where l_q is the duration of the given quantifier, and l_f is that of the given fragment. The same table presents the maximal contrasts in intensity and pitch within the corresponding entire clauses, with the purpose of identifying the significant difference in the falling of the intensity and the pitch:

Table 6: *Results of the statistical tests for the whole set (n=41) and the subset (n=31)*

		n=41; p<0,05	n=41; p<0,01	n=31; p<0,05	n=31; p<0,01
$Q_\exists$	duration	False	False	False	False
	intensity	True	True	True	True
	f_0	True	True	False	False
$Q_\forall$	duration	False	True	False	False
	intensity	False	False	False	True
	f_0	True	True	False	True

The cells with value *False* in Table 6 show the advantageous cases in which the null hypothesis that the difference is occasional is to be rejected.

The duration tests differentiated the two readings of the quantifiers in both cases and in both sets of subjects. The intensity differences between the two kinds of *each*-quantifiers (see Figure 1) also proved to be significant, at least with the weaker margin of error, that is, with 95% confidence (NB: as little as a 3 dB difference means a twofold intensity of the sound). It is also

presented in Figure 2 that there is a difference between the pitch values of the two *each*-quantifiers, but, according to Table 6, it only became significant after the exclusion of the problematic subjects. In the case of the subset (n=31), the two readings of the *also*-quantifier also proved to be significantly different in respect of pitch values.

It can be concluded that the two relevant-set based operators under investigation can also be produced and interpreted in appropriate situations as a topic.

4. Conclusion and loose ends

As a potential development of the experimental part, we plan to use multimodal stimuli in future researches: namely, visual stimuli (Surányi and Turi, 2017; Tóth and Csatár, 2017), or voice stimuli (Alberti and Szeteli, 2017), or the combination of the two.

On the one hand, the use of the research material of this paper as a further voice stimulus (Alberti and Szeteli, 2017) can serve as valuable control in new tasks for new subjects to measure the same factors. On the other hand, multimodal stimuli will make a better understanding of the test situation possible.

It is also an important research task, with the ultimate purpose of obtaining a comprehensive picture of the relevant-set based operators, to reconcile our research, in a theoretical as well as in a methodological sense, with such inspiring experiments concerning scope taking properties of quantifiers in Hungarian as Surányi and Turi's (2016; 2017) approach.

We conclude the paper with an analysis of chains of *wh*-words, which also serve as evidence for the existence of the two types of quantifiers.

É. Kiss (1992) claims that pre-V *wh*-word chains function according to the following formula of operators: Q*F. That is, only the last *wh*-word functions as an interrogative operator while the first k–1 *wh*-words substitute for ("disguised") universal quantifiers (10a), as corroborated by the congruent scheme of answer presented in (10a').

(10) a. Ki kinek \KIT mutatott be?
who who.Dat who.Acc introduce.Past in?
'Who introduced whom to whom?'
That is, 'Tell me, please, for *each* pair x and y of persons, who has been introduced to y by x.'

a'. Ali például Balázsnak és Csabának a \HÚgát mutatta be, Daninak és Edének pedig a \NŐvérét. Balázs Alinak egy \VOLT \BArátnőjét ajánlotta,...
'Ali, for instance, introduced his little sister to Balázs and Csaba while his elder sister to Dani and Ede. Balázs introduced to Ali one of his ex-girl-friends.'

b. (?)\KI kinek kit mutatott be?
who who.Dat who.Acc introduce.Past in?
'Who introduced whom to whom?'
'That is, 'I know that x introduced z to y. Identify, please, this triplet <x,y,z>.'

b'. Ali mutatta be Balázsnak a \HÚgát.
'Ali introduced his little sister to Balázs.'

c. ??Ki \KINEK kit mutatott be?
who who.Dat who.Acc introduce.Past in?
'Who introduced whom to whom?'
That is, 'I know that everyone introduced someone to someone. Identify, please, these pairs <y,z> introduced to each other for each "introducer" x.'

c'. Ali mutatta \BAlázsnak mutatta be a \HÚgát. Balázs \CIlinek a \BÁTYját. ...
'Ali introduced his little sister to Balázs. As for the pair introduced to each other by Balázs, he introduced his elder brother to Cili. ...'

Laczkó (2017, p. 215) quotes Mycock's (2010) pitchtrack (with *kinek* furnished with the stress pattern typical of the left edge of the predicate/comment zone):

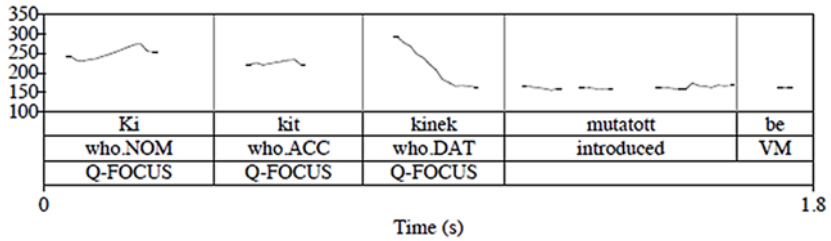


Figure 5: *Mycock's (2010) pitchtrack cited by Laczkó (2017)*

This prosodic pattern corroborates an analysis according to which the first two *wh*-words are topics in the topic–predicate/comment dimension (which "collect participants in order to anchor the comment to them") and *each*-quantifiers in the relevant-set based logico-pragmatic tier (all potential pairs of the relevant persons are referred to, as if the quantifier *mindenki* 'everyone' stood there), while the third *wh*-word functions as a real interrogative focus ("which persons satisfy certain criteria?"). The annotations 'Q-FOCUS' under all the three *wh*-words in Mycock's table shown above come from the automatic application of the traditional stance according to which a quantifier can be interpreted as anything else but a focus if, and only if, it serves as a contrastive topic. The take-home-message of this paper is that there are also non-contrastive-topic-like quantifiers.

Our approach predicts that the given *wh*-word chain can, indeed, be interpreted in a way that the *wh*-words all function as foci. We claim that it is possible to perform the chain in this way, as illustrated in (10b), and the resulting meaning is indeed the one predicted: a single triplet of persons is asked for. The congruent answer in (10b') corroborates this reading.

The examples in (10c-c') illustrate that, at least as a theoretical possibility, even the combination of one topic with two foci does exist.

References

Alberti, G., & Medve, A. (2000). Focus Constructions and the "Scope-inversion Puzzle" in Hungarian. In Gábor Alberti & István Kenesei (Eds.) *Approaches to Hungarian 7*. (pp. 93–117). Hungary, Szeged: JATEPress.

Alberti, G., & Szeteli, A., (2017). Szemafor hát! Manuscript, submitted to *Jelentés és nyelvhasználat* [Meaning and Language Use].

Blomgren, M., Chen, Y., Ng, M. L. & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America, 103,* 2649–2658.

Bródy, M. (1991). Remarks on the Order of Elements in the Hungarian Focus Field. In István Kenesei (Ed.) *Approaches to Hungarian 3.* (pp. 95–121). Hungary, Szeged: JATEPress.

Büring, D. (1997). *The Meaning of Topic and Focus. The 59th Street Bridge Accent*. London/New York: Routledge.

Büring, D. (2003). On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy*, 26, 511–545.

É. Kiss, K. (1992). Az egyszerű mondat szerkezete. [Structure of the simple sentence] In Ferenc Kiefer (Ed.) *Strukturális magyar nyelvtan 1. Mondattan.* [Structural grammar of Hungarian 1. Syntax] (pp. 79–179). Hungary, Budapest: Akadémiai Kiadó.

É. Kiss, K. (1998). Identificational focus versus information focus. *Language,* 74, 245–273.

É. Kiss, K. (2002). *The Syntax of Hungarian*. UK, Cambridge: Cambridge University Press.

Farkas, J., & Alberti, G. (2017). The inverse-scope puzzle of contrastive topics in Hungarian, which does not exist at all. Manuscript, submitted to *Approaches to Hungarian* 16.

Gécseg, Zs. (2013). Adalékok az információs szerkezet és mondatszerkezet összefüggéseinek kérdéséhez [On the issue of the interaction of information structure and sentence structure]. *Magyar Nyelv* [Hungarian Language] 109(1), 32–49.

Gerőcs, M., Babarczy A. & Surányi. B. (2014). Exhaustivity in Focus: Experimental Evidence from Hungarian. In *Language Use and Linguistic Structure. Olomouc Modern Language Series. 3.* 181–194.

Gyuris, B. (2009). *The Semantics and Pragmatics of the Contrastive Topic in Hungarian*. Hungary, Budapest: Lexica.

Hollien, H. & Wendahl, R. W. (1968). Perceptual study of vocal fry. *Journal of the Acoustical Society of America, 43,* 506–509.

Laczkó, T. (2017). *Aspects of Hungarian Syntax from a Lexicalist Perspective*. Hungary, Debrecen: DSc dissertation.

Markó, A. (2013). *Az irreguláris zönge funkciói a magyar beszédben* [The functions of irregular voicing in Hungarian speech]. Budapest, Hungary: ELTE Eötvös Kiadó.

Matteson, S. E., Olness, G. S. & Caplow, N. J. (2013). Toward a quantitative account of pitch distribution in spontaneous narrative: Method and validation. *The Journal of the Acoustical Society of America, 5,* 2953–2971.

McGlone, R. E. (1967). Air flow during vocal fry phonation. *Journal of Speech, Language and Hearing Research, 10,* 299–304.

Mycock, L. (2010). Prominence in Hungarian: The prosody-syntax connection. *Transactions of the Philological Society,* 108(3), 265-297.

Szabolcsi, A. (1997). Strategies for Scope Taking. In Anna Szabolcsi (Ed.) *Ways of Scope Taking.* (pp. 109–154). The Netherlands, Dordrecht: Kluwer.

Surányi, B. (2009). "Incorporated" locative adverbials in Hungarian. In Katalin É. Kiss (Ed.) *Adverbs and Adverbial Adjuncts at the Interfaces*. (pp. 39–74). Germany, Berlin: Mouton.

Surányi, B., & Turi, G. (2016). Az információs szerkezet szerepe a kvantorok hatókör-értelmezésében [The role of information structure in explanation of scope interpretation]. In Bence Kas (Ed.) *"Szavad ne feledd!" Tanulmányok Bánréti Zoltán tiszteletére* [Studies in honor of Zoltán Bánréti]. (pp. 233–246). Hungary, Budapest: Hungarian Academy of Sciences, Research Institute for Linguistics.

Surányi, B., & Turi, G. (2017). Focus and quantifier scope: An experimental study of Hungarian. In Harry van der Hulst & Anikó Lipták (Eds.) *Approaches to Hungarian 15*. (pp. 209–238). Amsterdam: John Benjamins.

Szűcs, P. (2017). *On Clause-Initial Discourse-Related Constructions in English and Hungarian*. Hungary, Debrecen: PhD dissertation.

Tóth, E., & Csatár, P. (2017). Preverbal Focus and Syntactically Unmarked Focus: A Comparison. In Joseph Emonds & Markéta Janebová (Eds.) *Language Use and Linguistic Structure, Proceedings of the Olomouc Linguistics Colloquium 2016.* (pp. 227–244). Olomouc: Palacký University.

Uriagereka, J. (1999). Multiple Spell-Out. In Samuel D. Epstein & Norbert Hornstein (Eds.) *Working Minimalism.* (pp. 251-282). Cambridge, MA.: MIT Press.

Varga, L. (2016). The Intonation of Topic and Comment in the Hungarian Declarative Sentence. *Finno-Ugric Languages and Linguistics,* 5/2, 46–77.

LingBaW *Linguistics Beyond And Within 4 (2018), 173-181*

Code switching and the so-called "assimilation narrative"

Tamás Vraukó
University of Miskolc, Hungary

Abstract

In literary theory, the works of (ethnic) minority authors–and similarly, the works of authors dealing with minorities–are often referred to as "assimilation narrative." This term tends to suggest that minority authors, who write in the language of their country, seek a place in society through assimilation. Assimilation, however, means melting up in the majority nation by adopting all the values, customs and way of life characteristic of the majority, and abandoning, leaving behind, giving up the original traditional values, ethics, lifestyle, religion etc. of the minority. Assimilation means disappearing without a trace, continuing life as a new person, with new values, language, a whole set of new cultural assets. In this paper an effort is made to show that this is in fact *not* what many of the ethnic minority writers look for, so the term *assimilation narrative* is in many, although certainly not all, the cases, erroneuosly applied. It is justified to make a distinction between *assimilation* and *integration* narratives, as the two are not the same. In the paper examples are provided from Hispanic-American literature (Mexican-American, Puerto Rican and Dominican), across a range of genres from prose through drama to poetry, and also, examples are discussed when the author does in fact seek assimilation, as well as stories in which neither assimilation, nor integration is successful.

Keywords: narrative, assimilation, Hispanic-American literature

1. Introduction

Assimilation and integration have long been burning issues on the European continent, where millions of ethnic minority people live in states other than their mother country, and recent developments, concerning the migration of what now appears to be millions of people into Europe, intensifies the debates. The United States has been proud of assimilating the crowds of people coming from Europe, Asia, Africa, Central- and South America. Today, however, it is more and more evident that what was often referred to as a "melting pot," is a "salad bowl" instead. In a salad bowl the ingredients are mixed up, but clearly identifyable. Interestingly, an Italian educational website uses the U. S. as an example of the melting pot that assimilates people into uniformity, and the U. K. as the example of the salad bowl, "with people of different cultures living in harmony, like the lettuce, tomatoes and carrots in a salad" (cglearn.it). A

Norwegian website now refers to the U. S. as a salad bowl: "The term complementary identity is frequently used to characterize the immigrant possessing both an ethnic identity and a national identity as an American citizen. To explain this double identity we often use "salad bowl" as a metaphor. In the "salad bowl" metaphor each culture retains its own distinct qualities (the different ingredients in the salad), but has a sense of common national identity in the country of habitat (the salad)." (ndla.no)

These are but two examples to illustrate how perplexing the relevant concepts and defitions are. The terms "assimilation" and "integration" are often used as close synonyms; something is absorbed, made similar to something else. Often not even the best sources of definitions provide clearcut guidelines. Oxford Advanced Learner's Dictionary of Current English defines integration as follows: "to become or make sb become fully a member of a community, rather than remaining in a separate group, e. g. because of one's race, colour etc." (Hornby, 1996, p. 620). The dictionary defines integration as a perfect synonym of assimilation. On the other hand, Webster's New World Dictionary offers the following definition of integration: "the bringing of different racial or ethnic groups into free and equal association" (Webster's 1986, p. 732). For this paper, the definition of Webster's dictionary is accepted, that favours *free* and *equal* association, and contains no reference to full cultural-ethnic assimilation.

What a lot of minority authors–prose writers as well as playwrights and poets–seek and expect from the majority society is a chance for *integration*. Immigrants coming to a new country leave their old homeland behind because they seek a better life, better financial and business opportunities, or they want to escape from political or religious harassment. None of these reasons necessitates the complete abandonement of centuries-old religious, social, ethnic, dietary, clothing, building, etc. customs and traditions. From the point of view of the immigrants, integration is a considerably more positive term, as it suggests the possibility for active participation in the life of society. It does not involve the abandonement of their original cultural heritage, but offers the advantages of participating in the life of their new country. The inhabitants of a territory occupied by another country will, naturally, see even fewer reasons for giving up their own civilization.

Linguistic arguments are (also) available to underpin the statement, and they, first and foremost, include vocabulary. Minority authors take pleasure in using terms of their native language in the text written for, and in the language of, the majority society. The "foreign" words serve to create an atmosphere of the culture of the source language for the readers in the majority society. At the same time, for many authors, using these words is a way of upholding their own cultural-ethnic identity in an environment ethno-culturally alien to them. Examples are taken from Hispanic-American prose, poetry and plays.

Hispanic-American literature was chosen because Hispanics–and. among them, Chicanos–constitute the largest ethnic minority in the U. S., and some of the conclusions may apply to other minorities. The period (cca. 1965-2000) was chosen because Hispanic literature became well-known and powerfully present in the U. S. with the advent of Anaya's *Bless me, Última!* in 1973, after the emergence of the Chicano Movement in the mid-1960 that drew public attention to the situation of the Latinos. No particular selection criteria were applied; the works were chosen because of their position in the literary canon. A brief look at any of the works discussed here tells about the attitude of their author to their own ethnic background, and to the issue of

assimilation or integration. The time frame was introduced because the first great wave of Hispanic literary works accomplished their mission, drawing public attention to the life and problems–often the mere presence–of the Hispanic community in the U. S. Anaya and other authors wrote other works following their first ones, but they are not always received as favourably as the original ones. This is the case with Anaya; the only exception is perhaps Luis Rodriguez, whose sequel to the first part of his autbiographical story is praised by critics. No ideological-political poems like *I am Joaquín* are needed now, and Hispanic authors seek new topics.

The choice of the Spanish terms largely depends upon the personal taste and preference of the individual authors, but usually the things and concepts they find especially important are written in Spanish, and the ones with which they wish to use to attain a certain artistic or, sometimes, political goal.

2. Code switching in prose

Rudolfo Anaya's *Bless me, Última!* (1973) was the first major Hispanic-American novel that drew the attention of the public to the life of the Hispanic minority. The very title suggest that Anaya is interested in his own people. Última is a *curandera*, a wise old woman, a person in possession of the ability to cure. She uses herbs, potions, and a little magic. Anaya's novel was published in 1972, when the Chicano Movement, started in the 1960s, powerfully asserted their political, civil rights and, last not least, cultural goals. Anaya in this novel introduced regular code-switching as a means of giving his story a cultural angle and keeping the reader aware of the cultural background in which the story takes place. The benevolant witchcraft of the *curandera* is a symbol, suggesting that Christian religion, the Catholic faith of the Mexican-Americans can be brought into harmony with the ancient beliefs of the Aztecs.

Another major Hispanic-American novel, *Always Running–La Vida Loca–Gang Days in Los Angeles* (1993) by Luis Rodriguez came two decades after Anaya's work. Here again, the title itself contains code switching. The setting is different, not the pastoral idyll of the open meadows, but the suffocating world of the *barrio.* The *barrio* is the Spanish for district, and is universally applied to the parts in cities populated primarily by Latinos. A lot of people live a life far from normal and desirable; instead they have to live *la vida loca,* a crazy life in the *barrio*. It is not easy to break out of the vicious circle of violence, poverty, deprivation and inadequate education. For Rodriguez's protagonist, art and literature offered the chance. He went to the library, and under the suspicious and contemptuous eyes of the librarian he selected books for himself. The author describes this revelation as follows:

> 'And then there was Piri Thomas, a Puerto Rican brother, *un camarada de aquellas:* his book *Down These Mean Streets* became a living Bible for me. I dog-eared it, wrote in it, copied whole passages so I wouldn't forget their texture, the passion, this searing work of a street dude and hype in Spanish Harlem–a barrio boy like me, on the other side of America.'
>
> (Rodriguez, 1993, p. 138).

Barrio is so natural a word that the author does not even italicize it. He does italicize *una camarada de aquellas*, that is, "a comrade from the streets," partly to indicate their common

Hispanic background, and partly because of the importance of his discovery. Then their similar social status is emphasized in English and half-English and half-Spanish: "street dude" and "barrio boy". (p. 138).

The first edition of the "living Bible" for Rodriguez's hero (in fact, Rodriguez himself) came out in 1967 and is, similarly to *Always Running,* largely autobiographical. Thomas wrote the novel as an adult, but when he was a young boy, he knew little about the dynamics of society, multiculturalism, assimilation, integration and such. The dark-skinned Puerto Rican boy wanted to be like the white men. To young Piri imitating the whites seemed to be the way of elevating his social status. Piri uses code switching when something important takes place in the story: for example, he wants to sleep with a *muchacha blanca*, that is, a white girl (Thomas, 1997, p. 93). Not because he could not have girlfriends, black or hispanic, but he definitely wanted a white girl. When he finally met one who was willing to sleep with him, he was so disgusted by the whites that he did not enjoy it al all. In the end, Catholicism and his experiences in prison bring him a revelation similar to what Rodriguez experience. He realized that being normal was cheap and easy, as compared to the life of a drug addict: "I made mental figures and my junkie *panín* needed seventy-two dollars a day to keep from coming apart–to stay normal. Something I was doing for nothing." (Thomas, 1997, p. 328). *Panín* means close friend, an intimate pal, and the Spanish term here indicates the importance Thomas ascribed to his relationship to fellow "street dudes".

Code switching is now a universal means of creating a special atmosphere in literature. An important, and similarly autobiographical novel is Julia Alvarez's *How the Garcia Girls Lost their Accents* (originally published in 1991). Alvarez is from Dominica, and fled to the United States from the paranoid dictator, Trujillo. She had lived an upper-middle class life in Dominica, they had automobiles, a swimming pool, exclusive parties, fancy clothes, and a lot of other amenities that most people did not have in the U. S. in the 1950s, let alone other parts of the world. The title suggests an assimilation narrative, and despite the fact that the Garcia girls settled in the U. S., had families of their own, they had lost their accent a lot before that, due to the excellent education they had been provided by their parents. When they return to Dominica for a visit after the fall of Trujillo, they immediately begin to look for their roots, their traditional drinks, music, although they are not disposing with what they acquired in America. They are successfully *integrated*, rather than *assimilated*, into American life. They pereserve a lot of the traditional Hispanic way of life, family values and merits: "These Latin women, even when the bullets are flying and the bombs are falling, they want to make sure you have a full stomach, your shirt is ironed, your handkerchief is fresh." (Alvarez, 2004, p. 197). Alvarez herself refutes the idea of assimilation with her own life, as she shares her time between her new home in New York, and Dominica, where she teaches local people to the methods of modern farming. Code switching is common all through the novel: when Yolanda returns to Dominica, she stops at a roadside *cantina,* and asks for *guava* juice. The *cantina* owner is unable to serve her that; instead, she proudly offers Coca-Cola, indicating how well supplied her *cantina* is. (Alv arez, 2004, p. 15).

Not everybody is, however, against assimilation. Sandra Cisneros in her *The House on Mango Street* (originally published in 1984), makes the American dream a strictly personal one, excluding ethno-cultural issues, and family as such. Unlike Thomas and Rodriguez, Cisneros

does not identify with the people living in difficulty; she wants to leave the whole misery behind once and for all. Her protagonists live in the *barrio,* but the term is avoided all through the short novel, although we even learn its exact geographical boundaries when somebody buys an automobile, and when they try it, they ride around the barrio.

Code switching is present at other places in the story: "The grandpa slept on the living room couch and snored through his teeth. His feet fat and doughy like tamales, and he powdered and stuffed into white socks and brown leather shoes." (Cisneros, 1991, p. 29) The author does not italicize tamales in the text when she uses the popular dish as a simile to present what she believes is a not very attractive image, grandfather's swollen legs stuffed into the socks and shoes, like the meat is stuffed into the *tamales*. In the case of Cisneros, the melting pot does not fail to melt; it functions well. The desires of the author are strictly personal, she does not mind melting up in American society, leaving behind her cultural and ethnic heritage. This is pure assimilation: personal goals, personal desires to be satisfied. Luis wanted to leave poverty and deprivation, Cisneros' protagonist, Esperanza, wanted to leave behind her entire former life, including her heritage, with all the negative and positive features and examples. After rejecting the *barrio* as a place to live in, she rejects the food of her *raza*, that is, her people, as well.

When the individual is unable to find a point where he can join the mainstream society, and he is unable either to be assimilated or integrated, the result is tragedy. Such a story is written by Richard Dokey, whose short story titled *Sánchez* came out in 1967. All the characters in it embody basically positive attributes. Juan Sánchez is a hard-working migrant, who makes a home in the United States. He marries the girl he loves, and who "accepted his philosophy completely, understood his needs, made it her own." (Dokey, 1991, p. 262). The loss of this beloved wife is the only tragedy in the story. Jesús, Sánchez's son, also works hard, finds a job among the Gringos, and is proud to show his place of work to his father. Jesús also shows his father the entertainment facilities near the place he works. This is the moment when Sánchez sadly realizes that he has lost his son. Jesús loves and respects his father with a true filial love, but he is absorbed by the world of the *norteamericanos.* So much so, that he does not even understand why his father is not equally enthusiastic about the wonderful things he shows him. After the loss of his wife, Sánchez has to release his son. The end of this short story is suicide–Sánchez sets all his belongings and finally his house on fire. The reason for his suicide is not some guilt, but his sense of unbearable loss. The values of the *norteamericanos*–a (relatively) well-paid job, entertainment, a new place in a competitive society–absorb a second-generation Chicano who is ready to adapt to this world. Sánchez went to the United States to find a better living. He worked hard, but he never really assimilated–he either did not want to, or was not able to, and remained a *paisano,* a man of simple needs, living close to the land he cultivated. Working hard is a value Sánchez shares with the *norteamericanos*–the only value, in fact, the only degree to which he was willing to integrate. Technical and social progress, entertainment, the company of people of the same age are not things he is looking for, and when he realizes that he has missed something, it is too late. He could have chosen either of two alternatives: showing more flexibility in connection with mainstream America, in other words, showing more inclination to assimilate, at least to some extent. The other is taking more care in bringing up his son in the traditions of his own nation, to keep him closer to his own *raza,* thus slowing him down on his way towards assimilation. But his son starts to assimilate at a pace that makes

it hopeless for Sánchez to catch up with. Sánchez did not simply perish in the fire, as the people in town believed: "But of course, on that score they were mistaken. Juan Sánchez had simply gone home." (Dokey, 1991, p. 267). Sánchez chose the way of his departure, having completed his mission on Earth.

Juan Sánchez is not an educated man, he is a farmer, not a man of letters, but when he uses code switching in the short story, and he does, it is always some positive, uplifting term. He refers to his wife as *bellissima*, that is, the most beautiful one. Sánchez appreciates the savage beauty of the Sierra where he lives.

3. Code switching in poetry

Poetry has always been an important "weapon" in revolutions and social movements, and that is the case with the Chicano Movement, the movement of the Mexican-Americans to better their life and to assert their social, ethnic and political goals. The long poem titled *I am Joaquin* by Rodolfo Gonzalez is often quoted for its political, rather than poetic merits. The opening of the poem is a powerful example of code-switching:

> Yo soy Joaquín,
> perdido en un mundo de confusión:
> I am Joaquín, lost in a world of confusion,
> caught up in the whirl of a gringo society,
> confused by the rules, scorned by attitudes,
> suppressed by manipulation, and destroyed by modern society.
> My fathers have lost the economic battle
> and won the struggle of cultural survival.
> And now! I must choose between the paradox of
> victory of the spirit, despite physical hunger

The first Spanish lines are repeated in English, showing that the poem was written for *both* audiences: Anglo- and Hispanic American. There is nothing about assimilation in it; the poem is a celebration of the Mexicans' dual heritage: European and Aztec. The finishing lines leave no doubt about the opinion of the poet:

> I am Joaquín.
> The odds are great
> But my spirit is strong,
> My faith unbreakable,
> My blood is pure.
> I am Aztec prince and Christian Christ.
> I SHALL ENDURE!
> I WILL ENDURE!

For the prominent poetess, Gloria Anzaldúa, code switching appears to be more natural than writing in just one language. That is how she recommends her book in the foreword of *Borderlands/La Frontera*: "This book is dedicated a todos mexicanos on both sides of the border." (Anzaldua, 1987). Spanish terms are natural, organic parts of the text, so no

italicization is necessary. In her poem *To Live in the Borderlands* she still italicizes Spanish terms:

> To live in the borderlands means you
> are neither *hispana india negra espanola*
> *ni gabacha, eres mestiza*, *mulata*, half-breed
> caught in the crossfire between camps
> while carrying all five races on your back
> not knowing which side to turn to, run from
>
> (Anzaldua, 1987, p. 194)

The poetess gives vent to the same problems as Rodolfo Gonzalez and, smilarly to his method, repeats some of the Spanish lines in English.

4. Code switching in drama

Hispanic theatre has deep roots in American soil. The first thing that Don Juan de Oñate and his expedition did when they arrived at the Rio Grande was stage a play. Campesino Theatre, that is, the "Peasant Theatre," was popular with the farmworkers movement and the Chicano Movement, growing out of that. Luis Valdez (re)invented the genre, offering support to El Movimiento from the stage.

Carlos Morton, Chicano playwright in his *Brown Buffalo* uses frequent code switching, and identifies his *raza* as a new one. Morton, who was born in the north, in Chicago, makes it clear that he does not want to see his ethnic community as an assimilated one:

> Well, the **Abuelos** want brown **Raza** (people) like me to create havoc, to stir the winds of change. We're the bronze race, you see, and Tonatiuh is the Sun God. Bronze is the color of the sun, the color of the earth, as in **La Raza**. (The race). (Morton, p. 19. Manuscript, courtesy of the author)

Similarly to the two poets quoted above, and a great many other Hispanic authors, Morton emphasizes the dual–European and American–heritage of the Hispanic Americans. In Morton's play, the protagonists desperately seek a position in American society, but they are certainly not satisfied with the role and position of the underdog. They are not free of sins, and they are even ready to embark on criminal activities, such as drug trafficking, in order to earn the money they need to live at standards they long for. They attribute their disadvantageous situation, their inability to integrate, to their being *different*. That is what they, in turn, take pride in: they are the *bronze race*.

Octavio Solis, contemporary playwright from California, uses code switching so frequently that it is not always possible to understand some of his texts for a reader/viewer who does not speak BOTH languages with some fluency. The example is from his "El Paso Blue":

Duane: Al, you owe me. We are camaradas, and as camaradas, we do for each other, we sacrifice.

> Al: Yo no te nebo ni madre, cabrón!
>
> (Solis, 1994, p. 93.)

5. Conclusions

Even this handful of examples illustrates that minority authors are not attracted by the idea of assimilation. They do want to be a part of a rich country, enjoy the advantages that it offers, and are ready to contribute to the joint efforts creating that welfare, but without giving up their ancient linguistic, religious, cooking, dressing etc. traditions. It was not easy in most cases; rejection by the majority of the society, initially inadequate education for the minorities, deep-rooted prejudices on both sides hampered the process for a long time. Luis Rodriguez, at the end of his novel, comes to the following conclusion: "It's about time we become part of America." (Rodriguez, 1993. p. 212). Rodriguez chooses his words carefully; instead of simply saying that it is time we became Americans, he talks about becoming *a part of America*. No assimilation is mentioned here, only integration. Robert N. Bellah's explanation and, at the same time, suggestion, is the following:

> What would you want me to tell my students about how they can fulfill their responsibilities as citizens?' one of us used to ask at the conclusion of his interview with community leaders. Almost always the characteristically American answer was 'Tell them to get involved!' The United States is a nation of joiners.
> (Bellah, 1996, p. 167)

Involvement means participation, fulfilling one's duties as a citizen, without ethnic-cultural amalgamation, without being lost in a uniform mass of people. In this way, Hispanic authors have been able to sustain their ethnic-cultural background, and in the 21st century they will no longer need to fight against assimilation with linguistic means or other. Movie makers also reach back to the works that are now regarded as classics: Luis Valdez' play, *Zoot Suit*, was converted into a major musical by Universal Artists in 1981, and Anaya's novel, *Bless me Última*, into a movie drama in Gran Productions in 2013.

References

Alvarez, Julia. (2004). *How the Garcia Girls Lost their Accent.* London: Bloomsbury.

Anaya, Rudolfo. (1994). *Bless Me, Ultima.* New York: Time Warner Books.

Anzaldua, Gloria (1987). *Borderlands/La Frontera.* San Francisco, Aunt Lute Books.

Barak, Julie. (1998). "Turning and Turning in the Widening Gyre": *A Second Coming into Language in Julia Alvarez' How the García Girls Lost Their Accent.* MELUS, Vol. 23, No. 1, Latino/a Literature (Spring, 1998),

Bellah, Robert N. et al. (1996). *Habits of the Heart. Individualism and Commitment in American Life.* Berkeley and Los Angeles: University of California Press.

https://www.cglearn.it/mysite/civilization/american-culture/a-story-of-immigration/melting-pot-or-salad-bowl/ (accessed 18 November 2018, 10:11)

Cisneros, Sandra. (1991). *The House on Mango Street.* New York: Vintage Books, Inc.

Richard Dokey. (1971). "Sánchez." *The Chicano–from Caricature to Self-Portrait.* (pp. 254-268). New York: New American Library.

Frever, Trinna S. "Oh! You Beautiful Doll" Icon, Image and Culture in Works by Alvarez, Cisneros and Morrison." *Tulsa Studies in Women's Literature.* Vol. 28. No. 1. Spring 2009.

Gonzalez, Rodolfo. *I am Yoaquín.* Retrieved from http://www.latinamericanstudies.org/latinos/joaquin.htm (accessed 11:48 a. m. 10 10 2017)

Guralnik, David B. [ed.] (1986) *Webster's New World Dictionary of the American Language.* New York: Prentice Hall Press

Hing, Bill Ong. (1997). To Be an American: Cultural Pluralism and the Rhetoric of Assimilation. New York: New York University Press.

Hornby, A. S. (1995) *Oxford Advanced Dictionary of Current English.* 5th edition, Oxford: OUP

Morton, Carlos. (1995). *The Brown Buffalo*. The Long version. Manuscript: courtesy of the author.

https://ndla.no/en/node/15153?fag=42 (accessed on 16 November 2018, 10:59)

Rodriguez, Luis. (1993). Always Running. La Vida Loca. Gang Days in L. A. New York: Touchstone.

Shirley, Carl R. and Paula W. Shirley. (1998). *Understanding Chicano Literature.* University of South Columbia: Carolina Press.

Solis, Octavio. (1994) "El Paso Blue." *Plays by Octavio Solis.* New York: Broadway Play Publishing Inc., undated

Thomas, Piri. (1997). *Down these Mean Streets.* New York: Vintage Books, 1997.