LingBaW

Linguistics Beyond and Within

VOLUME 6 (2020)

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

OPEN ACCESS

Articles

Fortis-lenis vs voiced-voiceless plosives in Welsh

Sabine Asmus
University of Szczecin, Poland
University of Leipzig, Germany

Sylwester Jaworski
University of Szczecin, Poland

Michał Baran
University of Szczecin, Poland

Abstract

This paper questions the voiceless-voiced distinction of Welsh consonants and claims that the fortis-lenis distinction is more appropriate for the description of the language. In light of research results of theoretical as well as experimental investigations into Welsh, e.g. the vowel-coda length dependence discovered by Asmus and Grawunder (2017), advocated further research into that matter, seeing also that the fortis-lenis distinction establishes a firm link to focal properties of Welsh, such as morpheme-initial consonant mutations (mICM). It was, therefore, decided to look at potential phonetic features that would contribute to the postulated distinction. These features are aspiration, voicing, hold phase duration and the centre of gravity (abbreviated to CoG) in the articulation of Welsh plosives. Preliminary results of the study discussed in this paper were summarised in "Fortis-lenis or Voiced-voiceless – features of Welsh consonants" (Asmus *et al.* 2019). However, expanding our research has yielded more comprehensive findings. As a result, it appears that the two series of plosives under review are different in terms of all features studied, but it is aspiration that is of major importance (thus confirming classifications of Welsh as an aspiration language).

Keywords: fortis-lenis distinction, plosives, consonant distinctions, aspiration

1. Introduction

Consonant distinctions are vital in describing a language because they facilitate any subsequent research of more complex language structures. Therefore, a language cannot be effectively analysed, taught or studied unless correct consonant distinctions are established. As far as Welsh is concerned, it is claimed here that its description as a fortis-lenis language is more appropriate than that as a voiced-voiceless tongue.

Some major accounts of the Welsh language (Ball and Fife 2002; Hannahs 2013) postulate that Welsh consonants are organised into a voiced and voiceless set. Such claims may result from the fact that the voiced-voiceless distinction is often seen as universal.[1] Ideal cases of languages based on the voiced-voiceless distinction should exhibit a phonological system, in which all obstruent phonemes can be arranged into voiced-voiceless pairs. The sound systems of Ukrainian (Shevelov 1996), Belorussian (Mayo 1996) or Polish[2] (Gussmann 2007; Sawicka 1995) may be quoted as examples of such consonantal inventories. For Welsh, however, this is certainly not the case as becomes clear in the course of this paper (see also Asmus *et al.* 2019 and Grawunder and Asmus 2017).

As Welsh is not the only language that does not feature such a phonological system, the universal status of the voiced-voiceless distinction has been challenged by various researchers. Some postulate that it is only one of the possible options of laryngeal contrast. Kiparski (2006), for instance, discusses so-called voicing and non-voicing languages, where the latter exhibit a different type of laryngeal contrast, not based on the presence or absence of phonetic voicing. Honeybone's (2005) laryngeal realism points to two major systems based on laryngeal contrast, one of which is the fortis-lenis distinction.[3]

Back in 1983, Jaeger discussed both notions referring to the terms 'fortis/lenis' as "used to characterize a basic phonological contrast in consonant systems which cannot be explained in terms of a voicing distinction" (1983: 177).[4] In context of the Insular Celtic languages, the voiced-voiceless distinction has also been questioned. A look at Bednarska (2016) reveals a fortis-lenis divide at work in Breton, which, like Welsh, is an Insular p-Celtic language.

Ball (1984), Jones (1984), Ball and Williams (2001), Morris and Hejná (2019) and Iosad (forthcoming) have also employed the fortis-lenis terminology in analyses, but rather in order to describe some features of Welsh consonants, e.g. aspiration where phonetic voicing is not enough to describe them (see Jaeger and the case of English below), than identifying Welsh as

---

[1] See Maddieson (2013) for such an approach. On the basis of such approaches and the features of the degree of obstruction of the vocal tract during the production of the sound and the absence or presence of voicing, a sonority scale was defined, which is by some also seen as universal (see Carr 2008: 160, Davenport and Hannahs 2005: 75).

[2] In standard Polish, the voiced velar fricative [ɣ] seems to have lost its phonemic status; however, it is still used in the south and east of Poland.

[3] It has also been established that languages may have a three-way distinction among fortis, lenis and aspirated plosives. In such systems, several articulatory, acoustic and aerodynamic factors contribute simultaneously to the contrast formation. According to Chang (2007), the lenis, fortis and aspirated plosives of Korean differ from each other with respect to linguo-palatal contact, glottal configuration, subglottal and intraoral pressure, laryngeal and supralaryngeal articulatory tension, voice onset time (VOT), fundamental frequency (f0) of vowel onset, intensity of vowel onset, and voice quality of vowel onset. Chang (2007: 21) also stresses the fact that "none of these cues alone differentiates all three series from each other due to a high degree of overlap between two or sometimes all three categories with respect to their range of realizations of these phonetic dimensions". Welsh distinguishes between a fortis, a lenis and an aspirated [r] (Jaworski and Asmus 2018), as well as between a fortis, lenis and aspirated variant of [l]. These may seem like an instance of a division similar to the Korean contrast, but they are in fact a remnant of an incomplete development of the four-fold Old Irish sonorant system in Welsh (see Asmus and Grawunder 2017), which came to an end in the 10th century CE, probably due to a limited influence of palatalisation resulting from language contact.

[4] Other uses of the term 'fortis/lenis' may refer to secondary phonetic features as is seen in the following.

a language that displays a fortis-lenis phonological system.[5] However, van Sluis described in 2019 that aspiration, together with general articulation length, is decisive for the distinction between the radical fortis plosives, lenited fortis plosives and radical lenis plosives in Old and Middle Welsh (2019: 75).

When Baran and Asmus (2019) revisited morpheme-initial consonant mutations (mICM) in Welsh and Irish, they showed that lenition processes, i.e. deaspiration /p, t, k/ > /b, d, g/, spirant mutation /p, t, k/ > /f, θ, χ/ (i.e. lenition of fortis plosives plus aspiration), approximantisation /ɬ, r̥ʰ/ > /l, r/ and debucculisation /s/ > /h/[6] established a whole system of phonetic contrasts that are used in order to mark morphological and syntactic features in the language. In addition, Asmus and Grawunder (2017) found out that there is little or no final devoicing – also referred to as contrast neutralisation of voiced and devoiced codas – in Welsh monosyllables ending in simplex codas. To the contrary, the maintenance of the fortis-lenis codas is semantically relevant and, whereas devoicing is largely absent, consonant length is decisive.

As can be seen, there is a clear phonological contrast in Welsh consonants, which cannot be explained by a voicing distinction. This is also confirmed by minimal pairs, which clearly follow the fortis-lenis divide. Whereas English features the voiced-voiceless[7] distinction between /s/ and /z/, like in *seal* and *zeal*, Welsh exhibits minimal pairs along the fortis-lenis divide[8] as can be seen in the following:

(1)   *sil* 'seed' vs *hil* 'breed'

In addition, this minimal pair[9] cannot be explained within the voiced-voiceless framework because both sounds are voiceless.

---

[5] It may be interesting though that Ball and Müller deny the applicability of this sonority scale to the Insular Celtic languages in 2016 and 2019.

[6] The contrast between /s/ and /h/ stopped being productive in Welsh in the 6th century CE and nowadays it is no longer productive in the mutation system (Jackson 1953: 12–513, 517–521, 525–527, 625), so it is often omitted by researchers (see for instance Ball 1984) but see also below.

[7] Some scholars refer to English in certain frameworks as a fortis-lenis language, the implication being that fortis sounds involve a greater amount of articulatory effort (Cruttenden 1996: 31). However, Jaeger rightly declines this idea (1983: 177–179) claiming that in English the voiceless stops are considered fortis only because of the presence of aspiration and the fortis-lenis distinction, understood in that way does not have any further phonological or phonetic implications. In the English language, the phonological status of aspiration is thus somewhat ambiguous due to it occurring in predictable contexts.

[8] This phonetic change /s/ - /h/ has already been referred to as a specific form of lenition, i.e. debucculisation, above. It is part of the lenition processes in Europe apparent in the 1st century CE. Whereas – as said above – lenition became grammaticalised in the Insular Celtic languages, it can still be found as a dialectal marker in several varieties of Spanish, e.g. Andalusian, Mexican or Caribbean Spanish (see Hualde 2005; Penny 2000).

[9] The word *sil* and *hil* can still be understood as 'seed' (for the confirmation thereof, we would like to thank Dr. Guto Rhys), although *hil* would normally be used in the sense of 'lineage' and 'race' these days (*Geiriadur Prifysgol Ar Lein*). Another good minimal pair is *sedd - hedd*. Both share the same etymology, but specialised on the current meanings 'seat' and 'peace' respectively. A shared etymology can also be established for *soch*/*swch* and *hwch* 'pig'. Although the latter is now a specific pig, i.e. 'a swine', the first is only loosely linked to the animal these days and normally understood as 'ploughshare' or 'snout'. However, the children's book *Jaci Soch* (Jones 2004) talks about the adventures of a little pig, not least because "*soch, soch*" may be used onomatopoetically in

In order to present the phonological fortis-lenis divide as a fully distinct one, phonetically distinctive features that distinguish between the two series must be identified. It can be assumed that features or their combination may be different for different manners of articulation. Therefore, the influence of voicing, aspiration, articulatory timing/hold phase duration and the centre of gravity (CoG) of the aspiration noise that follows the release of Welsh plosives is investigated here.

The remaining part of the article is structured as follows. Section 2 introduces the phonetic properties of plosives and describes how they contribute to making speech sounds distinct. Section 3 is devoted to an investigation into specific features of Welsh plosives, which might prove our claim, and specifies the objectives, describes the research methodology and presents the results. The article ends with conclusions, which also include suggestions for further research in this area.

2. Phonetic properties of plosives and their importance in laryngeal contrasts

The first of the features analysed in our current studies is phonetic voicing. Trask (1996: 381) defines it as vocal folds vibration happening while articulating a given sound. He claims that this may be the one parameter that distinguishes between voiced and voiceless sounds. Ladefoged (1975) in his feature system differentiates between five different values of the voice feature: glottal stop, laryngealised, voice, murmur and voiceless. Measuring voicing in Welsh plosives may answer the question whether this feature alone is enough to distinguish between the minimal pairs of sounds resulting from laryngeal contrast. If it were, it would provide phonetic evidence for a phonological voiced-voiceless distinction, but previous phonetic analyses of Welsh consonants such as Ball (1984) and Asmus and Grawunder (2017) suggest that the influence of phonetic voicing, especially word-finally, is incidental at the most.

The second feature, i.e. aspiration may generally be defined as 'a period of voiceless breathing', that follows an obstruent (Stevens 1998: 451; Trask 1996: 36). As far as experimental phonetics is concerned, aspiration manifests itself as friction following the release of a plosive. It is closely connected with the voice-onset time (abbreviated to VOT), i.e. the time between the release of the plosive to the beginning of modal voicing signifying a vowel. Ladefoged and Maddieson (1996: 45) argue that languages make use of three modal possibilities, namely, voiced, voiceless unaspirated and voiceless aspirated. When VOT equals zero or is very short, it normally means that a sound is an unaspirated voiceless plosive. Long, positive VOT may indicate an aspirated voiceless plosive and a negative VOT suggests that voicing starts before the release of the closure, a feature characteristic of voiced unaspirated plosives.[10] Aspiration,

---

Welsh for the sound of a pig (I would like to thank Prifardd Meirion MacIntyre Huws for this information). More examples could be listed for Welsh (see also Asmus and Grawunder at https://www.academia.edu/33509077/Language_structuring_consonant_mutation_s_in_Welsh_and_Irish). However, the /s/ - /h/ contrast is fully phonologically productive in other Insular Celtic language, i.e. Irish. All in all, it makes good sense to include this contrast in our investigation.

[10] In some language, e.g. Polish, phrase or sentence initial voiced plosives are always pre-voiced, i.e. vocal folds vibration begins during the hold phase, a few tens of milliseconds before the closure is released (Sobkowiak 2004).

together with general articulation length, is seen to have been decisive for the distinction between the radical fortis plosives, lenited fortis plosives and radical lenis plosives observable in Old and Middle Welsh before the lenited fortis plosives blended with the radical lenis ones forming the consonant system known from Modern Welsh (van Sluis 2019: 75).

The plosive sounds of the Welsh language, despite being assigned to either the lenis or fortis category, do not fall neatly into the pattern as they are aspirated, irrespective of the group they belong to. As a consequence, the fortis-lenis dichotomy in Welsh appears to be a combination of two factors, namely, aspiration and voicing. As for the former, the amount of aspiration following the Welsh plosives, as in other languages, e.g. English, is closely related to their place of articulation, with the velars being followed by the longest and the bilabials by the shortest period of aspiration. The latter feature is superimposed on aspiration to a varying degree, with the lenis plosives having typically more voicing than the fortis ones.

To illustrate the point, a lenis plosive, pronounced in the word *dŵr* 'water', is depicted in Figure 1. The second half of the 25-millisecond aspiration period is voiced as indicated by the pulses of vocal fold activity. In the case of /b/ and /d/, it is not uncommon for voicing to be present not only throughout the aspiration, but also during the preceding hold phase.
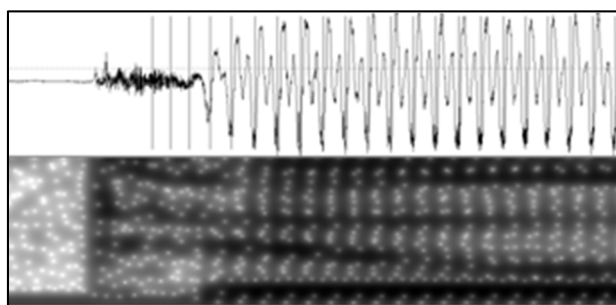
Figure 1: Partially voiced period of aspiration in the word dŵr 'water'

By contrast, fortis plosives are characterised by considerably longer aspiration periods, which tend to be voiced to a much lesser extent than those of lenis plosives. As illustrated in Figure 2, the period of aspiration of the /t/ plosive is 186 ms long, the last 17 ms of which is voiced. Although the voiced section of aspiration is slightly longer than that of *dŵr* (cf. Figure 1), it constitutes less than 10% of the total duration of the aspiration period. Since the Welsh lenis and fortis plosives differ with respect to the amount of aspiration they are followed by, as well as in regard to the duration of the voiced section of the aspiration noise, it is hypothesised in this work that both aspiration and voicing constitute crucial features underlying the fortis and lenis series.

The third factor under review is the hold phase, also called approach. It is sometimes claimed that in citation forms or after a pause, /p, t, k/ feature a longer hold phase than /b, d, g/, but the differences are not evident in connected speech (Ogden 2009: 98). Hold phase duration is here seen as one of the articulatory timing features and as such seems to be worth measuring because of the apparent vowel-coda duration dependence discovered by Asmus and Grawunder (2017). Recording statistically relevant differences in hold phase duration between the fortis and lenis series could suggest that length and articulatory timing are decisive in the fortis-lenis distinction.
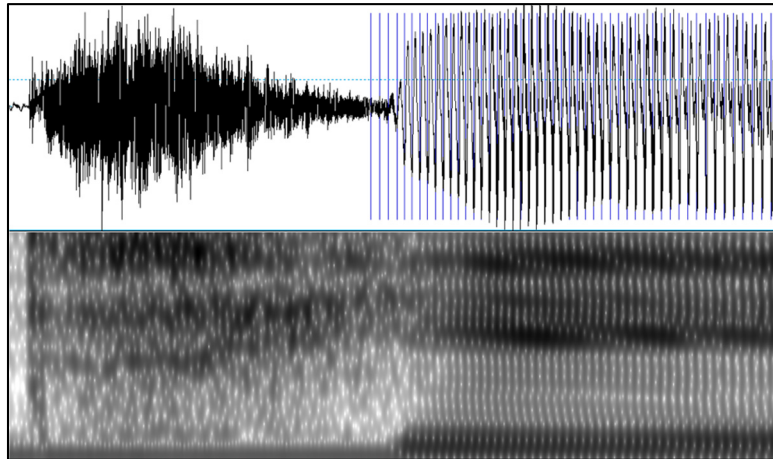
Figure 2: *Partially voiced period of aspiration in the word tir 'land'*

The last phonetic factor under review is centre of gravity of aspiration noise (henceforth CoG). This parameter can be thought of as the average frequency of aperiodic noise in a spectrum, which is correlated with the place of articulation of a fricative consonant (Ladefoged and Maddieson 1996; Stevens 1998). In general terms, the energy of front fricatives tends to be concentrated at higher frequencies than that of back ones. The burst release of a plosive consists of a certain amount of noise, which can be analysed in the same way as fricative sounds. If the CoG values of the aspiration noise following the Welsh plosives turn out to be significant, they might be considered as a potential phonetic factor that distinguishes between the lenis and fortis plosives.

3. The study

3.1. The objectives

As mentioned above, the objective of this study is to determine whether the four phonetic factors: aspiration, voicing, articulatory timing/hold phase duration and CoG contribute to the phonological fortis-lenis distinction in the case of the Welsh plosives. If so, the results will indicate that the voiced-voiceless distinction is inappropriate for the Welsh plosives and that the fortis-lenis divide appears to be phonologically distinctive and language-structuring.

3.2. The participants and methodology

In order to achieve the objectives of the study, we conducted an acoustic analysis of the Welsh plosives, i.e. /p, b, t, d, k, g/ produced in a reading task by 31 native speakers, male and female aged 19–71, from North and South Wales, who use their language at home and at work. The analysis was undertaken both in the onset and coda of monosyllabic native lexemes currently in use. The lexemes were taken from previous corpora of native vocabulary (Asmus and Grawunder 2017), but further amended by adding additional lexical items. An occasional English loan word was used as a control item. The tokens were then placed in the carrier phrase

Dw i heb ddweud X ond Y! [I didn't say X but Y][11] and were ordered in a way that every lexeme appears in the recording twice in a strong and twice in a weak prosodic position, with X being weak and Y being strong.[12] Selected native lexemes were also recorded in randomly chosen short sentences used as distractors. By so doing, we managed to record the same number of tokens of each plosive occurring in the same phonological context, which makes statistical analyses more reliable. The target items were subsequently examined with the help of the Praat software (version 5.3.85).

As regards the exact physical properties of the target sounds, the following features were measured: (i) the hold phase of word-final plosives,[13] (ii) the duration of the aspiration phase, (iii) the duration of the voicing phase that overlaps aspiration, (iv) the duration of the voiced part of the hold phase and (v) centre of gravity for the friction following the release of a plosive. With respect to aspiration, we define it is a period of friction extending from the release of a plosive to the onset of modal voicing. In the case of Welsh lenis plosives, the presence of friction following release is an indicator of aspiration. It is also assumed that friction following an initial lenis plosive may be partially voiced. Therefore, both parameters were measured in order to establish whether their relative durations distinguish fortis plosives from their lenis counterparts.

The collected data were subjected to a statistical analysis by means of a mixed-design ANOVA, which takes into consideration the influence of random effects, i.e. factors in an experimental design that are selected from a large population of potential samples rather than deliberately arranged by the researcher. In the case of this study, the random effects include the speakers and the phonological contexts in which the target sounds occur.

3.3. Results

As assumed above, the results of the analyses strongly suggest that the fortis-lenis dichotomy in Welsh is a function of the aspiration-voicing combination. In word-initial position, the lenis plosives /b, d, g/ seem to follow the pattern found in other languages (Ladefoged and Maddieson 1996). In word-final position, however, where aspiration of lenis plosives is cross-linguistically uncommon, the same duration pattern was formed, with /g/ followed by the longest and /b/ by the shortest period of aspiration (see Table 1 and Table 2). Periods of aspiration in the Welsh initial lenis plosives tend to be partially voiced, as shown in Figure 1 above. In the case of /b/ and /d/, it is not uncommon for voicing to be present not only throughout the aspiration, but also during the preceding hold phase. [14]

---

[11] Literally, 'I am without say(ing) X, but Y'.

[12] Based on the analysis of their recordings, Grawunder, Asmus and Anderson (2015) found out that the second position in the carrier sentence attracts stronger prosodic prominence.

[13] The duration of the hold phase of word-initial plosives was not determined as the informants would frequently make a short pause before producing a target word, which made accurate measurements impossible.

[14] The lack of hold phase duration data for initial plosives is due to the fact that, in numerous cases, the participants hesitated before pronouncing a target word, which made it impossible to measure the length of the hold phase with an acceptable degree of accuracy.

Table 1: *Average durations (in ms) of hold phases (HP), periods of aspiration (A) and periods of voicing of lenis plosives /b, d, g/ in prosodically weak position*

	Weak initial		Weak final		
	A	V	HP	V	A
b	17.3	9.5	74.6	22.7	47.3
	(± 3.1)	(± 2.5)	(± 13.7)	(± 7.8)	(± 18.6)
d	21.8	13.7	83.4	29.4	56.6
	(± 4.64)	(± 3.1)	(± 14.2)	(± 6.3)	(± 17.1)
g	33.4	17.1	97.3	40.1	66.2
	(± 5.1)	(± 4.4)	(± 15.6)	(± 7.2)	(± 17.9)
p	90.3	14.2	157.3	24.6	82.4
	(± 16.6)	(± 2.2)	(± 20.2)	(± 5.9)	(± 19.4)
t	113.4	16.9	177.8	19.2	96.9
	(± 20.7)	(± 3.3)	(± 26.4)	(± 6.8)	(± 22.5)
k	138.6	13.8	195.7	17.3	120.1
	(± 25.8)	(± 3.1)	(± 34.4)	(± 2.7)	(± 28.8)

Table 2: *Average durations (in ms) of hold phases (HP), periods of aspiration (A) and periods of voicing of lenis plosives /b, d, g/ in prosodically strong position*

	Strong initial		Strong final		
	A	V	HP	V	A
b	19.9	11.3	96.2	21.2	66.8
	(± 3.1)	(± 2.8)	(± 28.9)	(± 5.3)	(± 25.4)
d	22.8	12.1	109.5	20.4	89.6
	(± 4.2)	(± 2.5)	(± 38.7)	(± 4.6)	(± 19.7)
g	33.4	13.8	123.3	22.1	95.9
	(± 3.7)	(± 3.6)	(± 42.2)	(± 3.9)	(± 22.3)
p	117.6	13.4	208.2	17.4	119.6
	(± 24.5)	(± 2.6)	(± 39.8)	(± 3.1)	(± 27.7)
t	129.3	12.2	203.6	13.4	124.7
	(± 34.6)	(± 2.1)	(± 46.3)	(± 2.4)	(± 39.4)
k	148.2	13.5	226.9	16.1	142.6
	(± 41.4)	(± 1.9)	(± 49.5)	(± 2.6)	(± 40.8)

There is speaker-specific variation regarding the amount of aspiration. Predictably, the speaker effect turned out to be significant ($p < .0001$). It is worth pointing out that the Welsh lenis plosives have considerably more aspiration than voiced plosives in non-aspiration languages, e.g. (Sobkowiak 2004), Polish, but also in English (Ladefoged and Maddieson 1996). The difference seems to relate to the amount of friction noise, which includes intensity and/or duration, following the release; an aspect that will be further investigated in a different study.

With respect to the final lenis plosives, they are regularly aspirated to a greater extent than initial ones. Importantly, the aspiration noise is never voiced, yet a certain amount of voicing may occur in the hold phase (cf. Asmus and Grawunder 2017). The amount of aspiration following final /b, d, g/ forms the same pattern as in initial position.

By comparison, the fortis plosives /p, t, k/ are characterised by significantly longer periods of aspiration and relatively shorter periods of voicing in the hold phase. In the case of the hold phase, the differences are highly significant, regardless of the place articulation of plosive

(p<.00001). The data in Table 1 and Table 2 indicate that this finding refers to both prosodic positions investigated in this study.

The analysis of the CoG values of the aspiration noise following the Welsh plosives produced inconclusive results. Cross-linguistically speaking, the CoG values, falling within the 2,400–3,700 Hz range, tend to be much lower than those obtained for sibilant fricatives produced in various places of articulation (Stevens 1998; Żygis *et al.* 2015). What is more, in the study conducted by Wallin and Koffi (2017), which focuses on whispered speech, the CoG values for the non-sibilant fricatives [f], [v], [θ], [ð] are much lower than those of the sibilants [s], [z], [ʃ], [ʒ]. Thus, the findings strongly suggest that the aspiration friction that follows the Welsh plosives is generated at the glottis.
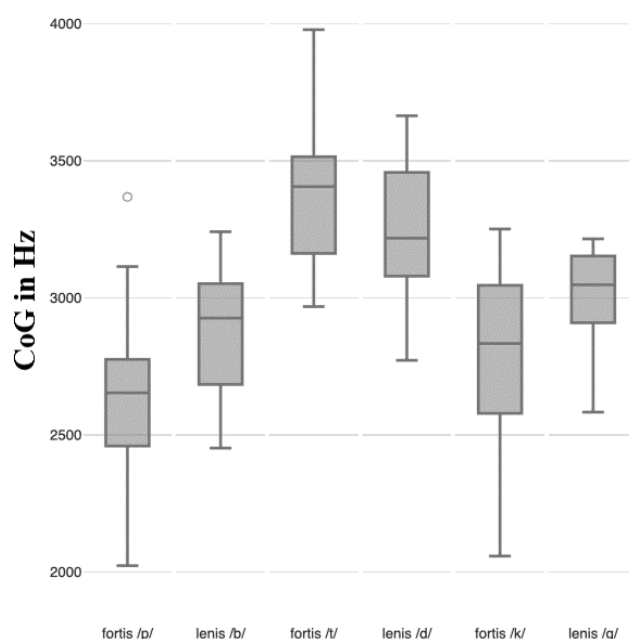
Figure 3: *Mean CoG values of aspiration noise following the Welsh plosives*

When the place of articulation is taken as a variable, the data presented in Figure 3 form a pattern inconsistent with what has been established regarding the relationship between the place of articulation of a plosive and its CoG. For both the lenis and fortis series, the mean CoG's of the aspiration following the velars are lower than in the case of alveolars, while the aspiration following the bilabials has the lowest CoG values. The differences between the velars and alveolars, as well bilabials and alveolars, turned out to be statistically significant (p<.05), whereas those between the velars and bilabials did not reach the level of statistical significance (p>.05).

Somewhat surprising are the data relating to the fortis-lenis pairs produced in the same place of articulation. The general impression is that the friction that follows the lenis plosives has a higher CoG than that of its fortis counterpart. In our study, this rule holds for the bilabials and velars, with statistically significant differences between the lenis and fortis segments. As for the alveolars, the CoG of the aspiration following /d/ is lower than that of /t/, yet the difference did not appear to be significant in terms of statistics (p=0.1259).

4. Conclusion

The authors of this paper claim that the fortis-lenis distinction is more appropriate to describe the phonological system of Welsh than the voiced-voiceless one, at least when looking at Welsh plosives. The claim is based on previous research, e.g. into basic phonetic processes at work in the Welsh mICM system, the correlation between the length of simplex coda consonants and preceding vowels in Welsh monosyllables and minimal pairs in the language, which cannot be explained by a voice-voiceless distinction.

This research was continued and supplemented by an investigation into specific features of Welsh consonants, which promised to prove our claim, because, ideally, the proposed consonant distinction should be identified by a set of phonetic correlates in order to be seen as separate from the voiced-voiceless distinction. Supported by a literature review, four phonetic correlates, namely, aspiration, voicing, hold phase duration and centre of gravity, were selected for this analysis.

The collected data point to the conclusion that both voicing and aspiration contribute to differentiating between fortis and lenis plosives. The period of aspiration following the lenis plosives is significantly shorter than that of their fortis counterparts in both word-initial and word-final position, while the differences in voicing do not form a consistent pattern. Also, the ratio between the length of the voiced period of aspiration and its total duration is significantly higher in the case of /b, d, g/. A similar relationship can also be established in word-final position, where the voiced part of the hold phase of the lenis plosives appears to be relatively longer than that of fortis /p, t, k/. Fortis plosives were found to have significantly longer hold phases than their lenis counterparts (see Tables 1 and 2). These findings agree to a considerable extent with data reported by Ball (1984), Jones (1984), Ball and Williams (2001), Morris and Hejná (2019) and Iosad (forthcoming). However, we think the data form a basis for seeing the fortis-lenis distinction from a wider perspective, not restricted to accounting for aspiration in plosives.

In brief, the acoustic investigation of the first two phonetic features, i.e. aspiration and voicing, offers promising phonetic evidence for a potentially phonological fortis-lenis divide as indicated by grammatical and semantic patterns. The effects of aspiration and voicing contribute to our claim that Welsh should be classified as a fortis-lenis language rather than a voiced-voiceless one, in particular when considering the general articulation length of Welsh simplex codas and the devoicing of lenis codas (Asmus and Grawunder 2017). What is more, general articulatory length (hold phase and aspiration) forms a pattern similar to that described by van Sluis (2019) for Old and Middle Welsh.

With regard to the CoG data collected for the purposes of the study, they seem to be inconclusive, yet they indicate that the aspiration noise following a plosive is generated at the glottis. Somewhat surprisingly, however, the fortis plosives, with the exception of /t/, are followed by friction whose CoG is lower than that of their lenis counterparts.[15] For this reason,

---

[15] In the study by Wallis and Koffi (2017), [v] has a higher CoG than [f], while the CoG of [θ] is higher than that of [ð]. In the case of sibilants, the value of CoG is always higher in the case of voiceless fricatives.

the data must be treated with caution and can only be regarded as one of the potential phonetic criteria that differentiate between lenis and fortis plosives.

References

Asmus, S., and S. Grawunder. 2018. *Language-structuring consonant mutation(s) in Welsh (and Irish).* Retrieved from https://www.academia.edu/33509077/Language_structuring_consonant_mutation_s_in_Welsh_and_Irish, Accessed at 2020-10-04)

Asmus, S., and S. Grawunder. 2017. *Vowel length in Welsh monosyllables, its interrelation with Irish and related problems.* Lewiston: Edwin Mellen Press.

Asmus, S., Jaworski, S., and M. Baran. 2019. Fortis-lenis or voiced-voiceless – Features of Welsh consonants. In S. Calhoun, P. Escudero, M. Tabain and P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 3773–3776. Canberra: Australasian Speech Science and Technology Association Inc.

Ball, M. J., and J. Fife. 2002. *The Celtic languages.* London: Routledge.

Ball, M. J., and N. Müller. 2016. Sonority and initial consonant mutation in the Celtic languages. In M. J. Ball, and N. Müller (eds.), *Challenging sonority. Cross-linguistic evidence*, 276–294. Sheffield: Equinox.

Ball, M. J., and N. Müller. 2019. Sonority and initial consonant mutation in Modern Celtic. In M. Bloch-Trojnar and M. Ó Fionnáin (eds.), *Centres and peripheries in Celtic linguistics*, 11–26. Bern: Peter Lang.

Ball, M. J. 1984. Phonetics for phonology. In M. J. Ball and G. E. Jones (eds.), *Welsh phonology. Selected readings,* 5–39. Cardiff: University of Wales Press.

Ball, M. J. and B. Williams. 2001. *Welsh phonetics.* Lewiston: The Edwin Mellen Press.

Baran, M., and S. Asmus. 2019. A new perspective on Welsh and Irish consonant mutations. In S. Asmus, and K. Jaworska-Biskup (eds), *New perspectives of Modern Wales – Studies in Welsh language, literature and social politics*, 157–177. Cambridge: Cambridge Scholars Publishing.

Bednarska, K. 2016. *Quantity in Breton and Welsh.* Lublin: KUL.

Carr, P. 2008. *A glossary of phonology.* Oxford: Blackwell.

Chang, Ch. 2007. Korean fricatives: Production, perception, and laryngeal typology. *UC Berkeley Phonology Lab Annual Report* (2007): 20–70.

Cruttenden, A. 1996. *Gimson's pronunciation of English.* Oxon: Routledge.

Davenport, M., and S. J. Hannahs. 2005. *Introducing phonetics and phonology.* Oxon: Routledge.

GPC Online. University of Wales Centre for Advanced Welsh and Celtic Studies. http://www.geiriadur.ac.uk.

Gussmann, E. 2007. *The Phonology of Polish.* Oxford: Oxford University Press.

Grawunder, S., Asmus, S., and C. Anderson. 2015. On the correlation of acoustic vowel and coda duration in Modern Welsh C(C)VC monosyllables. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, and J. Scobbie (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. London: The International Phonetic Association.

Hannahs, S. J. 2013. *The phonology of Welsh.* Oxford: Oxford University Press.

Honeybone, P. 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In M. van Oostendorp, and J. van de J. Weijer (eds.), *The internal organization of phonological segments*, 319–354. Berlin: Mouton de Gruyter.

Hualde, J. I. 2005. *The sounds of Spanish*. New York: Cambridge University Press.

Iosad, P. Forthcoming. The ATR/Laryngeal connection and emergent features. Submitted to B. Botma, and M. van Oostendorp (eds.), *Primitives of phonological structure.*

Jackson, K. H. 1953. *Language and history in Early Britain.* Harvard University Press.

Jaworski, S., and S. Asmus. 2018. An acoustic study of Welsh and Slavonic rhotics. *Studia Celto-Slavica* 9: 95–120.

Jaeger, J. J. 1983. The fortis/lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics* 11: 177–189.

Jones, G. E. 1984. The distinctive vowels and consonants of Welsh. In M. J. Ball and G. E. Jones (eds.), *Welsh phonology. Selected readings,* 40–64. Cardiff: University of Wales Press.

Jones, M.V. 2004. *Jaci Soch*. Aberystwyth: Cymdeithas Lyfrau Ceredigion.

Kiparski, P. 2006. The Amphichronic Program vs. Evolutionary Phonology. *Theoretical Linguistics* 32(2): 217–236.

Ladefoged, P. 1975. *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P., and I. Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell Publishers.

Maddieson, I. 2013. Voicing in plosives and fricatives. In M. S. Dryer, and M. Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info, Accessed on 2020-11-17.)

Mayo, P. J. 1996. *A grammar of Byelorussian.* Sheffield: University of Sheffield.

Morris, J., and M. Hejná. 2019. Pre-aspiration in Bethesda Welsh: A sociophonetic analysis. *Journal of the International Phonetic Association* 50(2): 1–25.

Ogden, R. 2009. *An introduction to English phonetics.* Edinburgh: Edinburgh University Press.

Penny, R. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press.

Shevelov, G. 1993. Ukrainian. In B. Comrie, and G. G. Corbett (eds.), *The Slavonic Languages*. 947–997. London, New York: Routledge.

Sawicka, I. 1995. Fonologia. *Gramatyka współczesnego języka polskiego*. Vol. 3. *Fonetyka i fonologia*. 105–195. [The grammar of the contemporary Polish language. Vol. 3. Phonetics and phonology.]

Sobkowiak, W. 2004. *English phonetics for Poles*. Poznań: Wydawnictwo Poznańskie.

Stevens, K. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

van Sluis, S. 2019. Lenited voiceless stops in Middle Welsh: Phonology and orthography. Ph.D. diss. Cardiff: University of Wales.

Trask, R. L. 1996. *A dictionary of phonetics and phonology.* Oxon: Routledge.

Wallin, J. and E. Koffi. 2017. The acoustic correlates of fricatives in whispered speech. *Linguistic Portfolios* 6: 101–108.

Żygis, M, Pape, D., Jesus, L., and M. Jaskuła. 2015. The effects of intonation on acoustic properties of fricatives. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, and J. Scobbie (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences.* 1–5. London: The International Phonetic Association.

Clitics in imperative: proclisis, enclisis and mesoclisis in Albanian and in Italo-Romance varieties of Lausberg area

Benedetta Baldi and Leonardo M. Savoia
University of Florence, Italy

Abstract

The relation between morpho-syntactic structure and its externalization into interpretive levels is the topic of this article. In many languages, typically in Romance and Albanian varieties, modal contexts, specifically imperative and infinitive, and negation, give rise to phenomena of clitic reordering and an interesting micro-variation. Imperative differs from declarative sentences in selecting enclisis except in negative contexts. Moreover, in Albanian mesoclisis appears in the 2nd plural person of imperative, between the verbal base and the person inflection. A similar distribution characterizes Calabro-Lucanian varieties spoken in Lausberg area, in contact with Arbëresh (Italo-Albanian) dialects. This article proposes to analyze the influence of modal contexts on the lexicalization of object clitics (OCls) and their different behavior in connection to their referential properties. Our descriptive and theoretical starting point is the representational morpho-syntactic approach adopted in Manzini and Savoia (2011 and subsequent works; see Section 5).

Keywords: imperative mood, mesoclisis, negative contexts, Albanian varieties, Romance varieties

1. Introduction

In this article, we will investigate some Albanian and Italo-Romance varieties that share mesoclisis in imperatives, i.e. the insertion of object clitics between the root and the inflection. The Albanian varieties include the Gheg spoken in Shkodër, the Tosk of Gjirokastër and the Arbëresh (Italo-Albanian) varieties of Firmo (Calabria), San Costantino Albanese (Basilicata), Greci (Campania) and San Marzano di San Giuseppe (Apulia). As to the Italo-Romance dialects, we have dealt with the imperative forms of Terranova Pollino and Senise (Basilicata), and Morano (Calabria), all belonging to the so-called Lausberg area, i.e. the conservative Romance area on the border between Basilicata and Calabria and surrounding the Pollino Massif, also including a number of Arbëresh villages in contact conditions. Finally, in order to substantiate our analysis, the doubling of object clitics in the North Apulian dialect of San Severo, and the imperative inflection in the South Calabrian dialect of Cardeto will be

examined. As is known from the literature, all of these varieties possess a system of object clitics that occur in proclisis in declarative sentences and in enclisis in imperatives, and, in some Romance languages, also in infinitive contexts. Here, we will analyze the distribution of OCls in imperative, as being a clue to understand its syntactic properties.

2. Imperative mood: A non-veridical form

Imperatives are strictly linked to a specific illocutionary force, an act of command (Aikhenvald 2010), addressed to the recipient. The reference to the latter provides the content for the interpretation of the subject. In other words, the subject of the imperatives is the recipient of the locutory production in the context of the speech act. So, the imperative generally excludes tense distinctions, insofar as it is deictically anchored to the discourse universe. Naturally, formal properties of the verb vary across languages, depending on their inflectional organization, so that in many languages imperative can coincide with the lexical base/root of the verb, whereas in others imperative is in turn endowed with person-specialized morphology (Aikhenvald 2010, Alcázar and Saltarelli 2014).

From a typological perspective, imperatives show the reversal of the more generally assumed markedness scale between 1st and 2nd person, in the sense that "third person is marked, while second person is often unmarked, particularly in the singular" (Alcázar and Saltarelli 2014: 41). Aikhenvald (2010: 76) schematizes the continuum from non-canonical (left) to canonical (right) 'person distinction' in imperative in the scale in (1):

(1)   1sg and/or 1p   >   3 sg or pl   >   1p inclusive   >   2p (sg, pl, or nonsingular)
      exclusive   non-singular

Thus, imperative overturns the markedness hierarchy underlying assertion/declarative sentences, in which 1st person subjects introduce the less marked type of references. This is evidenced by languages with subject clitics where it is the 1st person clitic pronoun that most frequently can be missing in the paradigm (cf. Manzini and Savoia 2005, forthcoming).

Both syntax and semantics of imperatives raised some debated issues. Platzack and Rosengren (1998: 192), assuming a cartographic approach, conclude that imperatives imply a sentence structure lacking the anchoring to the tense, mood and finiteness projections, usually characterizing declaratives and interrogatives, whereby "[t]he absence of FinP prevents the imperative clause both from referring to the event expressed and from containing a proper subject". The gist is that the order establishes a connection between the 'prominent argument of the sentence', the addressee to whom the order is given, and the event referred to by the proposition. Nevertheless, the non-finite nature of imperatives does not exempt them from presenting agreement properties independently of the modal and temporal specifications. All things considered, the crucial imperative property is the fact that the eventive content of the verbal lexical element is not associated to temporal or other referential morphosyntactic devices.

From a semantic point of view, imperatives introduce a clause that escapes the truth conditions generally applicable to declaratives (Giannakidou 1998). According to Portner

(2004) imperatives lexicalize a property rather than an event. In a nutshell, a very usual analysis is that the imperative form assigns a property to a prominent argument, identified with the addressee (Platzak and Rosengren 1998, Mauck and Zanuttini 2005). This identification is governed by the association via Merge (Mauck and Zanuttini 2005) of the prepositional argument with the Addressee selected by the Speech Act. This mechanism is explained in Zanuttini (2008: 196) by postulating that the imperative is the head of a JussiveP projection, which "has an operator in its specifier that... takes as input a proposition, consisting of the predicate saturated by the subject, and yields as output a property. This property has a presupposition that its argument, corresponding to the subject, refers to the addressee(s)." In other words, in the imperative a modal operator connects the situation denoted by the sentence in its scope to the hearer denotation.

Leaving out the structural solutions adopted in the literature, we agree with the proposal that the imperative denotes a property "which can only be true of the addressee" (Portner 2004: 239), rather than an event, and as such submitted to veridicality requirements. Schematizing, the idea of Mauck and Zanuttini (2005) is that imperative is a predicate with an unsaturated variable x bound by the λ operator introduced by the modal element as in (2), where *ʃix* is the 2nd person imperative form coinciding with the verbal root exemplified in (3a).

(2)   ʃix
       λx, see (x, y)

Naturally, the expressed (or covert) agreement of imperatives coincides with the features identifying the addressee – in many languages this is externalized by specialized inflectional exponents (Aikhenvald 2010, Alcázar and Saltarelli 2014), as in the case of the inflection of 2nd plural *-te* in Standard Italian and many Romance dialects, *-ni* in Albanian varieties. This raises an interesting issue, since the 2nd person inflection and object clitics lexicalize referential properties, possibly undergoing truth conditions, contrasting with the counterfactual nature of the imperative. This discrepancy can be seen as the reason for clitics in imperative clauses to be lexicalized by specialized forms in comparison with declarative ones (Manzini and Savoia 2017).

3.   Clitic distribution in imperatives: Albanian

The micro-variation among Tosk (Ghirokastër), Gheg (Shkodër) and Arbëresh varieties primarily involves the position of object clitics separating varieties allowing for total mesoclisis, as Albanian Tosk in (3), partial mesoclisis, as generally Arbëresh dialects of Italy in (4–5–6), or selectively preventing mesoclisis, as Shkodër Gheg in (7). Positive imperatives are exemplified in (a)–(a') for 2nd singular and (b)–(b') 2nd plural forms. (a)–(b) illustrate the insertion of a 3rd person clitic, (a')–(b') illustrate the insertion of the 1st person clitic, (c)–(c') exemplify the 1st person+3rd person accusative clusters and (d)–(d') the 3rd person dative +3rd person accusative clusters. As to the inflection of the imperative, it is of note that the 2nd singular person coincides with the root of the verb, whereas the 2nd plural person has the morpheme of 2nd plural *-ni*, as in *ʃixni!* 'see (2nd plural)!', *silni!* 'bring (2nd plural)!' (Gjirokastër) (for a comparison, see the negative forms in (8)). The 1st plural clitic is generally avoided or, eventually, it occurs before

the verb, although mesoclisis or enclisis seem to be accepted by some speakers, as exemplified in (e). The data in (f) illustrate the structures where the dative clitic *i* is inserted alone; in this variety it occurs in mesoclisis like the other 3rd person clitics. Finally, for the sake of completeness, we present OCls in pre-verbal position in declaratives, precisely accusatives in (g), dative in (g') and reflexive/unaccusative in (g") (cf. Manzini and Savoia 2007). It is of note that generally the 3rd singular person OCl is realized by the alternant *ɛ* in isolation and by the alternant *-a* in clusters *oblique-accusative*, as shown by the comparison between (3–7a,b) and (3–7c,d).

(3)   a.   ʃix- ɛ   / i
            see- him/her / them
            'see(sg) her/him/them'

      a'.   ʃix- mə
            see   me
            'see(sg) me'

      b.   ʃix- ɛ-   ni
            see- him/her 2PL
            'see(pl) him/her'

      b'.   ʃix- mə- ni
            see- me   2PL
            'see/dress(pl) me'

      c.   jɛp- i-   a
            give him/her it
            'give(sg) it to him/her'

      c'.   nə-/jɛp- m- a
            give   me it
            'give(sg) it to me'

      d.   jɛp- i-   a- ni
            give- him/her it 2PL
            'give(pl) it to him/her/them'

      d'.   nə-/jɛp- / sil   m- a / i-   ni
            give   / bring me it / them   2PL
            'give/bring(pl) it to me'

      e.   prit-   na- ni
            wait for us- 2PL
            'wait for(pl) us'

      f.   jɛp- i-   ni kətə
            give- him/her 2PL this
            'give(pl) him/her/them this'

      g.   mə / tə / ɛ   / i   / na / ju   θərasin
            me / you / her/him   / them   / us / you(pl)   they.call
            'they call me/you/her/him/them/us/you'

      g'.   i   a ðatʃ
            to him/her it I.gave
            'I gave it to him/her'

      g".   ai   u   vɛʃ
            he   refl he.dressed
            'he dressed up'

                                                                    Gjirokastër

Differently from Tosk varieties, in Arbëresh dialects, in the case of the clitic string *dative+accusative* deictic clitics (1ˢᵗ person) are incorporated inside the word, while the 3ʳᵈ person clitics are inserted on the right of the inflection. The following data come from the varieties of Firmo (Calabria) in (4) and San Costantino Albanese (Lucania) in (5). The simple occurrence of dative in (f), is obscured by the coalescence of the dative *-i* and the final vowel of the ending *-ni*, suggesting that its position is the same as the other 3ʳᵈ person elements.

(4)   a.   vɛʃ- ɛ   / sərrit- i
           dress him/her / call   them
           'dress(sg) her/him' / 'call them'

      a'. sərrit- im
           call   me
           'call(sg) me'

      b.   vɛʃ- ni- ɛ /   zɟɔ-   ni- ɛ
           dress 2PL him/her / wake up 2PL   him/her
           'dress(pl) him/her' / 'wake up(pl) her/him'

      b'. sərrit- m- (n)i / zɟɔ-   m- ni
           call   me 2PL / wake up me 2PL
           'call(pl) me'   / 'wake up(pl) me'

      c.   jip- i-   a
           give 3DAT it
           'give (sg) it to her/him'

      c'. ɔ-   m- ɛ
           give me it
           'give (sg) it to me'

      d.   jip- ni- i-   a
           give 2PL him/her it
           'give (pl) it to her/him'

      d'. ɔ́-   m- ni- ɛ
           give me 2PL it
           'give(pl) it to me'

      e.   zɟɔ-   n(i)- na
           wake up 2PL   us
           'wake(pl) us up'

      f.   jip- ni(i) kit   ʃurbɛs
           give 2PL this thing
           'give this thing to him/her'

      g.   mə / tə / ɛ   / i   / na / ju   ʃɔkin
           me / you / her/him / them / us / you(pl) they.see
           'they see me/you/her/him/them/us/you'

      g'. i   a japjin
           to him/her it they.give
           'they give it to him/her'

      g". ai u   uʎ
           he refl he.sat
           'he sat up'

                                                              Firmo

(5)   a.   siɛl- ɛ   / i
           bring him/her / them
           'bring(sg) her/him/them (towards the speaker)'

a'.   ciɛl- əm
     bring me
     'bring(sg) me (away from the speaker)'
b.   sil-   ni- ɛ
     bring 2PL him/her
     'bring(pl) her/him up'
b'.   cɛl-   m- i
     bring me 2PL
     'bring(pl) me!'
c.   siɛl- m- ɛ
     bring me it
     'bring(sg) it to me'
c'.   sil-   m- ni- ɛ
     bring me 2PL it
     'bring(pl) it to me'
d.   jip-   i-   a
     give   him/her it
     'give(2sg) it to him/her'
d'.   cɛl-   n- i-   a
     bring 2PL her/him it
     'bring(2pl) it to him/her'
f.   jip-   n (i/ -i)   ktə
     give- 2PL him/her this
     'give this to him/her'

<div align="right">San Costantino</div>

It is worth noting that in (5) the 2nd plural person of imperative presents two alternants *-i* and *-ni*, the first of which is the usual inflection of the 2nd plural person of indicative. In the imperative *-ni* becomes obligatory in mesoclisis contexts, where it is preceded by the deictic 1st person clitic.

   Some differences emerge in Arbëresh dialects. In *Greci*, the accusative clitic follows the imperative, as in (6a), whereas the 1st person clitic precedes it, as in (6b). When they combine, the 1st person clitic is in mesoclisis, while the accusative is inserted in enclisis.

(6)   a.   zɟɔ-   nni a
          wake.up 2PL him
          'wake(pl) him up'
     b.   mə / na zɟɔ-   nni
          me / us wake.up 2PL
          'wake me/us up'
     c.   ne- m-   ni a
          give to.me 2PL it
          'give it to me'

<div align="right">Greci</div>

In the Gheg variety spoken in Shkodër, the 1st person clitics and clitic clusters *1st person+3rd person* occur in pre-verbal position, in this differentiating from the preceding varieties, which insert 1st person clitics in enclisis or in mesoclisis. 3rd person clitics and 3rd person clusters (accusative and dative) occur in post-verbal position, or, variably, in mesoclisis. We note that in 2nd plural forms with the 3rd person cluster *j-a* '3dat+3acc' the duplication of the 2nd plural

inflection is attested, as in (7d). Moreover, in Shkodër variety the cluster *i-a* doubles a lexical object so that we do not find the lexicalization of the simple dative *i* (cf. (3d)).

(7)   a.   ʃif-   ɛ
           look at- him/her
           'look at(sg) her/him'

      a'. m   ʃif
           me look at
           'look at(sg) me'

      b.   θir- ni- ɛ
           call 2PL him/her
           'call(pl) her/him'

           ʃif-/θir-   ɛ-   ni
           look at/call   her/him   2PL
           'look at / call (pl) her/him'

      b'. m   ʃif-ni   / θir-ni
           me look-PL / call-pl
           'look at / call (pl) me'

      c.   nɛp-   i-   a
           give.2SG 3DAT it
           'give(sg) it to him/her'

      c'. m- a / na ɛ   nɛp
           me it / us   it   give
           'give(sg) it to me/us'

      d.   nɛp- ni- i-   a- (ni)   / nɛp- i-   a- ni
           give 2PL him/her it   2PL   / give- him/her   it   2PL
           'give(pl) it to him/her'

           tʃɔ- ni- i-   a- (ni)   (ktə)
           bring 2PL him/her it   2PL
           'bring(pl) it to him/her'

      d'. m   a   nɛp- / jɛp- ni
           1SG it   give-   2PL
           'give(pl) it to me'

      g.   m   / t   / ɛ   / i   / na / ju   ʃef
           me / you / her/him / them / us   / you(pl) (s)he.sees
           '(s)he sees me/you/her/him/them/us/you'

      g'. i   a   nɛp
           to him/her   it   (s)he.gives
           '(s)he gives it to him/her'

      g". u   lɑva
           refl I.wash
           'I wash up'

                                                                              Shkodër

What is to note is that the order between deictic and 3rd person clitics is substantially preserved as in other varieties, in the sense that deictic clitics occupy a position in a domain more to the left than the 3rd person clitics. This distribution is realized by placing deictic clitics before the verb and 3rd person clitics inside or to the right of the verbal form.

Negative imperatives involve a specialized negation *mɔs/mas* occurring in other modal contexts, and entail clitics to be inserted pre-verbally, in proclitic position, before all persons. The negation precedes the clitic string and the following verb, as illustrated in (8).

(8)   mɔs   i-   a   jɛp
        neg   him/her   it   give
        'do not give(sg) it to him'
        mɔs   m-   a   sil-   ni
        neg   me- it bring- 2PL
        'do not bring(pl) it to me'

                                                                           Gjirokastër

        mɔs   i-   a / m- ɛ   jip-ni
        neg   him/her   it / me it   give-2PL
        'do not give(pl) it to him/me'
        mɔs   na   zɟo-   ni
        neg   us   wake up 2PL
        'do not wake(pl) us up

                                                                            Firmo

        mɔs   m   ɛ   siɛl
        neg   me   it   bring-2SG
        'do not bring (sg) it to me'
        mɔs   m   ɛ   sil-ni/sil-i / jip-i
        neg   me   it   bring-2PL / give-2PL
        'do not bring/give(pl) it to me'

                                                            San Costantino

        mas   ɛ   / m   ʃif   / ʃif-ni
        neg   him/her / me look-2SG / look-2PL
        'do not look at (sg/pl) her/him/me'
        mas   m   / i   a nɛp   / nɛp-ni
        neg   me / him/her it give-2SG / PL
        'do not give(sg/pl) it to me / to her/him'

                                                               Shkodër

In the Arbëresh of San Marzano (Apulia) (cf. Manzini and Savoia 1999, 2007), otherwise showing the split between 1ˢᵗ person and 3ʳᵈ person clitics on a par with the other Arbëresh varieties, cf. (8a), in negative contexts the object clitic is doubled in pre-verbal and enclitic position, as in (9b).

(9)   a.   hua- nni- i-   a
               tell   2PL   him   it
               'tell(2pl) it to him/her'
               hua- mmə- ni- ɛ
               tell   me   2PL it
               'tell(2pl) it to me'
       b.   mɔsə m   ɛ   hua-nni- ɛ
               not   me   it   tell-2PL   it
               'do not tell it to me'
               mɔs i   a hua-nni- i- a
               not   him it tell-2PL   him it
               'do not tell(2pl) it to him'

                                                               San Marzano

Note that we have focused on the canonical 2nd person addressee forms of imperative (Kuryłowicz 1964), while we have left out the inclusive (you and I) 1st plural person. In Tosk and Gheg varieties, the 1st plural forms are realized by subjunctive forms, as in (10), introduced by the modal particle *tə* (MPrt), and the negation, in turn, implies the modal form *mɔs/mas*. In Arbëresh varieties we find the indicative form or the deontic periphrasis *have+inflected verb*, as illustrated in (10'); in the negative contexts the declarative contexts negation *nəŋg* 'not' occurs.

(10)   (mɔs)   t   a   vesh- im
     neg   MPrt   him/her dress 1PL
     'do not dress him/her'
     (mɔs)   t   i-   a jap- im
     neg   MPrt   him/her it give 1PL
     'do not give it to him/her'

<div align="right">Gjirokastër</div>

     (mas)   t   i-   a jap- im
     neg   MPrt   him/her it give 1PL
     'do not give it to him/her'
     (mas)   t   la-   hɛ-   na
     neg   MPrt   wash NA *Infl* 1PL
     'do not wash'

<div align="right">Shkodër</div>

(10') (nəŋg) (kɛ-   mi) i-   a jam- mi
     neg   (have 1PL) him/her it give 1PL
     'we (do not) have to give it to him/her'

<div align="right">Firmo</div>

     (mɔs)   i-   a jap-mi
     neg   him/her it give-1PL
     'do not give it to him/her'

<div align="right">San Costantino</div>

In Albanian, 1st person inclusive is realized by subjunctive, as in (10), whereas in Southern Italian dialects illustrated here, the speaker inclusive form is not differentiated, as in Standard Italian. We recall that in the varieties we investigate in this article, only 2nd person may select a specialized form, not necessarily coinciding with the root; in the other cases we find forms coinciding with the corresponding ones of present indicative or, in the case of 1st inclusive in Albanian, with subjunctive.

Resorting to subjunctive in non-canonical persons, specifically 1st plural, is a widespread strategy for imperative. The general difficulty or impossibility for languages to have imperatives addressed to 1st and 3rd persons is overcome by using counterfactual forms such as exhortative and subjunctive (Aikhenvald 2010, Alcázar and Saltarelli 2014). As we saw in Section 2, imperative introduces a speech act interpretation whereby the only admitted subject is the recipient of the interaction, typically the 2nd person addressee. This restriction is reflected not only interpretively, excluding self-ascription (1st person addressee), but also structurally, entailing asymmetry between speaker and recipient (Alcázar and Saltarelli 2014: 106).

3.1. Non-Active imperative forms

In non-active forms of imperative the middle-reflexive or passive interpretation is generally lexicalized by the non-active (NA) clitic *u* (*oneself*; Manzini and Savoia 2007), characterizing passive, middle and reflexive forms of the verb also in the perfect and, according to the different varieties, in other tenses. In the imperative, the position of *u* is similar to that of the 3rd person clitics. In fact, *u* is in enclisis in the 2nd singular imperative, as illustrated in (11a)–(11'a). In the 2nd plural person in (11b)–(11'b), it occurs in mesoclisis in Gjirokastër and in enclisis in Shkodër variety. The negative form in (11a',b')–(11'a',b') involves the reordering with the result that the modal negation *mɔs* precedes the sequence *clitic* u + *verb*. The data of Gjirokastër highlights the fact that differently from consonant roots, vocalic roots such as *la-* 'wash' select the middle-reflexive infix *-h-*.

(11)   a.   viʃ-   u
            dress   NA
            'dress(sg) yourself'
            la-   h-   u
            wash- NA *inflection*   NA
            'wash(sg) yourself'
       a'. mɔs   u   viʃ   / la
            neg   NA   dress   / wash
            'do not dress/wash (sg) yourself'
       b.   viʃ-   u-   ni
            dress   NA   2PL
            'dress(pl) yourself'
            la-   h-   u-   ni
            wash   NA *infl*   NA   2PL
            'wash(pl) yourself'
       b'. mɔs   u   viʃ-   ni   / la-   ni
            neg   NA   dress-   2PL   / wash   2PL
            'do not dress/wash (pl) yourself'

                                                                              Gjirokastër

(11') a.   lɑj-   u
            wash- NA
            'wash(sg) yourself'
       a'. mas   u   lɑj
            neg   NA   wash
            'do not wash(sg) yourself'
       b.   lɑ-   ni-   u
            wash- 2PL-   NA
            'wash(pl) yourself'
       b'. mas   u   lɑ-   ni
            neg   NA   wash   2PL
            'do not wash(pl) yourself'

                                                                              Shkodër

In Albanian varieties, the middle-reflexive morpheme *-h-ɛ* is inserted between the vocalic root and the inflection, while consonantal roots exclude *-h-* and generally select a specialized root

internal vocalic alternant. Summarizing, in 2nd plural middle-reflexive, imperatives in (11b) introduce the active inflection, while the middle-reflexive reading is lexicalized by the NA clitic, so excluding the middle-reflexive specialized infixes.

Differently from the Albanian varieties spoken in Balkan area, in Arbëresh dialects the internal structure of 2nd person plural of the middle-reflexive imperative illustrated in (12b) coincides with the 2nd person plural of the middle-reflexive indicative. More precisely, the vocalic root is followed by the middle-reflexive infix, -*h*- in San Costantino dialect, -*k-ɛ* in the one of Firmo, followed in turn by the person ending.

(12) a.   ʒɟɔj-   u
           wake up   NA
           'wake up(sg)'

      a'. mɔs   u   ʒɟɔ
           neg   NA   wake up
           'do not wake up(sg)'

      b.   ʒɟɔ-   k-ɛ-   ni
           wake up NA *Infl* 2PL
           'wake up(pl)'

      b'. mɔs   ʒɟɔ-   k-ɛ-   ni
           neg   wake up NA *Infl* 2PL
           'do not wake up(pl)'

                                                            Firmo

      a.   ʎaj-   u
           wash   NA
           'wash(sg) up'

      a'. mɔs   u   ʎaj
           neg   NA   wash
           'do not wash(sg) up'

      b.   ʎa-   h-   i
           wash NA *Infl* 2PL
           'wash(pl) up'

      b'. mɔs   ʎa-   h-   i
           neg   wash NA *Infl* 2PL
           'do not wash(pl) up'

                                                    San Costantino

In negative imperative the clitic *u* is reordered before the verb and preceded by the modal negation, as in (12a',b'), in the singular, while in negative 2nd plural the middle-reflexive indicative form is preserved, so that non-active interpretation is entirely devolved to the NA inflection -*h-, -k-*.

3.2. *An intermediate summary*

The phenomena we have described encompass the following points:
   – Mesoclisis is attested in 2nd person plural form of imperative in Arbëresh and in Gheg (Shkodër) and Tosk (Gjirokastër) varieties of Albania.
   – In Arbëresh varieties, mesoclisis affects only the 1st person singular form; 3rd person and NA clitics follow the imperative.

- In Gjirokastër variety, mesoclisis involves both 1st singular and 3rd person clitics.
- In Shkodër variety, 1st person clitics precede the imperative and mesoclisis involves only 3rd person and NA clitics.

Before addressing the interaction between the distribution of person elements, the different behavior of clitics and imperative form, some distributional phenomena concerning Southern Italian dialects will be considered. This comparison will shed light on the principles regulating the differential distribution of clitics and inflections in imperatives.

4. Mesoclisis and reordering in Lausberg area dialects

Calabro-Lucanian Lausberg area dialects are characterized by a distribution of clitics in imperative very similar to the one of contact Arbëresh varieties. In (13–14), data are provided from a Lucanian variety (Terranova Pollino) and a North Calabrian variety (Morano). In these dialects, unlike Albanian varieties, mesoclisis involves dative and locative clitics in addition to the 1st/2nd person forms and is restricted to the contexts where the 3rd person clitic occurs in enclisis. In other words, it affects only *deictic/dative/locative clitics + 3rd person clitics* clusters.[1] In (13a)–(14a) 2nd singular forms are provided, in (13b, b')–(14b) 1st plural forms and in (13c)–(14c) 2nd plural ones. At least in some of these dialects enclitic forms of clusters in 1st/2nd plural are variably realized, as exemplified in (13c'). (13d) and (14d) illustrate the postverbal position of the simple object clitics. Finally, in these dialects, as in general in Italian dialects, 1st plural imperative is normally attested and admits mesoclisis. The examples in (13e)–(14e) illustrate the object clitics occurring in proclisis to the verb in declarative sentences. These are realized by the simple inflectional elements *a, u, i* 'her, him, them', unlike the more complex structure of enclitic forms. We recall that in these dialects the clitic *nə* lexicalizes both partitive and dative, as in *n u ˈðɔnaðə* '(s)he gives it to him/her' (Terranova), *n u rɔŋgu* 'I give it to her/him' (Morano).[2]

(13) a.   ðɔn-a-   mˈm / nˈn -   illə/a:
           give-TV   me   / him/her   it
           'give it to me/him/her'
      b.   mənd-a-   ˈtʃ-   iə-m-   a:
           put- TV   LOC   TV-1PL   it
           'let us put it here'

---

    b'.   ðɔn-a-   nˈn-   iə-m-   a:
         give-TV   him/her   TV-1PL   it
         'let us give it to him/her'

    c.   ðɔn-a-   ˈm- iə-t-   a:
         give-TV   me   TV-2PL   it
         'give it to me'

    c'.   ða-   tə-   ˈm-   illə
         give- 2PL   me   it
         'give it to me'

    d.   ða-   tə-   nə   kwistə
         give- 2PL   him/her   this
         'give this to him/her'

    e.   u   / a   / i   viɣə
         him/it   / her   / them   I.see
         'I see him/it/her/them'

                                                  Terranova Pollino

(14)   a.   rɔn-a-   ˈmi / nˈni   lu
         give-TV   me   / him/her   it
         'give it to me/him/her'

    b.   rɔn-a-   nˈni-   mu-   lu
         give-TV   him/her   1PL   it
         'let us give it to him/her'

    c.   rɔn-a-   mˈmi- tu-   lu
         give-TV   me   2PL   it
         'give it to me'

    d.   caˈm-a-   mu-   lu
         call-TV   1PL-   him/her
         'let us call him/her'

    e.   u   / a   / i   vir-   unu
         him/it   / her   / them   see-   3PL
         'they see him/it/her/them'

                                                  Morano Calabro

As noted, the morphology of 3rd person OCls in enclisis includes the *l-* root characterizing D elements in Romance languages (Manzini and Savoia 2017). Terranova dialect in (13) shows also the morpho-phonological alternant *-a:* deriving from the velarization of original *l-*. In proclisis 3rd person clitics coincide with the bare gender/number inflection, as in (13e)–(14e).

    The enclitic elements attract the main stress of the word, as shown in the examples in (13–14). The prosodic reorganization triggered by the positioning of the main stress in the enclitic forms gives rise to a trochaic foot such as *dɔna-m ˈm-illə* 'give mi it' (Terranova Pollino), or an antepenultimate stressed string, such as *rɔna-mˈmi-tu-lu* 'give (pl) it to me' (Morano). In both contexts, it is the first clitic of the cluster that is lexically designated to attract the stress. The result is that a left-headed foot is created which is followed by a final reduced foot in 1st/2nd plural forms of imperatives, as in (15).

(15)   F   F
        | h |
        mmi tu   lu

Re-assignment of the main stress can be connected to the specialized nature of 3rd person clitics in post-verbal position in imperatives, where, following Manzini and Savoia (2017), the richer morphology is connected with the necessity to externalize a stronger referential import.

In the same way as in Albanian varieties, negative contexts imply the proclitic occurrence of pronominal forms, as exemplified in (16a-c) and (17a-c). Specifically, we find the 3rd person *l-* forms, in (16a) and (17a), i.e. the forms occurring in the position adjacent to the negative marker also in negative declaratives. As to the imperative form, in negative contexts the 2nd singular is lexicalized by the infinitive, exactly like Standard Italian, as in (16a) and (17a). In the other persons the usual person inflection occurs, as in (16b,c) and (17b,c), where tv stands for Thematic Vowel.

(16) a. ɔ-   ll-u   spətˈt-a-ðə
          neg   3-MSG   wait-TV-INF
          'do not wait for him'
          ˈɔ-   mm-u   / nn-   u   ðɐʊ-ðə
          neg   me-3.MSG / him/her-   3.MSG   give-INF
          'do not give it to me/him/her'
      b. ɔ-   nn-   u   dav-iə-mə
          neg   him/her- 3.MSG   give-TV-1PL
           'do not give it to him/her'
      c. ɔ-   mm-   u   dav-   iə-tə
          neg   me   3.MSG   give-   TV-2PL
          'do not give it to me'

<div align="right">Terranova Pollino</div>

(17) a. nu   ll-u: caˈm-ɛ
          neg   3-MSG   call-TV/INF
          'do not call him'
          nu   nn   u   ruˈn-ɛ
          neg   him/her 3.MSG   give-TV/INF
          'do not give it to him/her'
      b. nu-   nn-   u   ruˈn-ɛ-mu
          neg   him/her 3.MSG   give-TV-1PL
          'do not give it to him/her'
      c. nu-   mm-u: purˈt-ɛ-ti
          neg   me-3.MSG bring-TV-2PL
          'do not bring it to me'

<div align="right">Morano</div>

Summarizing, we see that:
  − Mesoclisis is triggered only when *dative/locative/1st person +3rd person clitics* clusters are inserted in 1st and 2nd plural forms of imperative, unlike Albanian varieties, where mesoclisis is not limited to the clitic clusters.
  − Dative/locative and 1st person clitics occur between root and inflection while accusatives occur on the right of inflection.
  − Negative imperatives are characterized by the same type of reordering as the one observed in Albanian varieties, whereby clitics and clitic clusters are inserted between the negation and the verb, in proclisis.

- In this dialects the enclitic form of accusatives is different from the proclitic elements, specifically including the definiteness lexical base *l-*, missing in proclitic elements.
- The enclitic forms attract the main stress of word as exemplified in (15).
- *l-* clitics occur also when preceded by the negation.

4.1. *Other cases of interaction between root, inflection and clitics in imperative*

Before concluding this section, we will extend the comparison to two other phenomena in which imperative imposes a special lexicalization of referential formatives, clitics or inflection. This requirement is clearly manifested by the occurrence of 1st person clitic or clitic clusters in Albanian varieties in (3–7) between the root and the inflection. What is more, some varieties provide evidence for duplication of the argumental positions, inflectional or clitics, externalized in two copies. This is the case of the reduplicative structures of Shkodër in (7d), e.g. *nɛp-ni-j-a-(ni)* 'give-2pl-him/her-it(-2pl), *give it to him/her*', where is the inflectional morpheme that occurs in a position adjacent to the root and in a position on the right. Similar duplicative structures are documented in Lucanian and Calabrian dialects with mesoclisis in (13–14) (Manzini and Savoia 2005, 2011). More precisely, we find both duplication of the inflection, as in the Shkodër examples, in (18a), and the duplication of the clitic, as in (18b).

(18) a. *duplicated inflection*
          pərtɔ-ma-   nˈn-   iəm-a:
          bring-1PL-   him/her-   1PL- it
          'bring it to him/her'

                                                                  Terranova Pollino

      b. *duplicated clitic*
          ra-   ˈm-itə-   mə- lə
          give-   me-2PL-   me- it
          'give it to me'

                                                                  Senise (Lucania)

A partially different condition is the one documented by San Costantino in (5), where the 2nd person plural of imperative introduces *-ni* in contexts with clitics, so that the simplex form *sil-i* 'bring (2pl)' alternants with *sil-m-ni-ɛ* 'bring-me-2pl-it, *bring it to me*'.

   Reduplication of clitics is independently attested in Apulian and Lucanian varieties. For instance, in the North-Apulian dialect of San Severo in monosyllabic 2nd singular person imperatives two copies of the first clitic of a cluster occur, as in (19a,a') (Manzini and Savoia 2005). (19b) illustrates the negative imperative with proclisis and (19c,c') the multisyllabic forms, excluding reduplication.

(19) a.   da-ttʃa-   ˈtʃ-illə
           give-us-   us- it
           'give it to us'
      a'. di-   mma-ˈm-illə   a mɛ
           tell- me- me-it   (to me)
           'tell it to me'

    b.   nə   mm u dɛ-nnə
          neg me   it give-GERUND
          'do not give it to me'

    c.   purta- ˈtʃ-illə
          bring- us- it
          'bring it to us'

    c'. di-ta- ˈm-illə
          tell-2PL- me- it
          'tell it to me'

                                     San Severo (Apulia)

As we see, a morpho-phonological restriction limits reduplication to the combination *monosyllabic root+clitic cluster*. However, the point is that the duplication of the clitic object is made possible by the imperative, suggesting that imperative introduces a domain on its right able to host object clitics.

    As the last case study, we consider the extension of the infinitival inflection *-ri* to the 1st and 2nd plural inflected forms in negative imperatives in some South Calabrian dialects. As we saw in Section 4, negative imperatives realize the 2nd person singular by means of the verbal infinitive. In the dialect of Cardeto (and other neighbouring villages in South Calabria, cf. Ledgeway *et al.* 2016) *-ri* occurs in the 2nd singular, in (20b), and is extended also to the other forms, as in (20b') for 1st plural person and (20b") for 2nd plural person. (20a,a',a") illustrate positive imperatival forms of 2nd singular, 1st plural and 2nd plural.

(20) a.   ˈcama- lu   / la   / mi
           call-   him / her / me
           'call him/her/me'

      a'. da-   mu- ntʃi-llu
           give- 1PL- him-it
           'let us give it to him/her'

      a". cama-ti-lu   / mi
           call-2PL-him / me
           'call him/me'

      b.   nə   mmi cama-ri
           neg me   call-INF
           'do not call me'

      b'. nɔn tʃ-u   da-mu-ri
           neg   him/her-it give-1PL-INF
           'do not give it to him/her'

      b". nɔn cama-ti-ri a   nnuḍu
           neg call-2PL-INF PREP no one
           'do not call anyone'

                                     Cardeto

Again, the imperative domain seems to be able to include a richer inflectional structure than the other verbal forms, admitting a duplicated insertion of inflectional material. The specialized occurrence of *-ri* in Cardeto and adjacent dialects has been addressed in Loporcaro (1995), Ledgeway *et al.* (2016), de Angelis (online). While keeping to different theoretical approaches, the authors assume that the insertion of *-ri* in the negative imperatives is due to an analogical leveling whereby the negative forms have re-established the parallel between the 2nd singular

person and the other ones. According to Ledgeway *et al.* (2016) this could derive from the contact with Greco imperative paradigm, where also the plural has dedicated forms. Loporcaro (1995) connects the leveling with the pattern of the positive form, that includes only two distinct morphologies (see the discussion in de Angelis online). What we can conclude is that the extension of *-ri* maybe depends on an analogical process, although its cause does not seem clear.

5. The structure of lexical elements and the theoretical framework

The asymmetries highlighted by the different distribution of clitics in positive and negative imperatives put into play some of the crucial morpho-syntactic properties of the sentence: the Phase structure of the imperative, the referential properties of the pronominal elements and the interpretive nature of the negation:

- DOM effects emerge, whereby deictic clitics (1st person) have a different distribution from 3rd person clitics at least in Shkodër and Arbëresh varieties.
- The 1st person clitics precede the 3rd person clitics in all contexts (mesoclisis/post-verbal/pre-verbal).
- Negation requires clitics to occur in pre-verbal position (Manzini and Savoia 2007, 2011, 2017; Baldi and Savoia 2018)

Let us begin by considering the differential distribution of clitics. The table (19) schematizes the distribution of object clitics in the 2nd person plural in (3–7) and in (13–14). In (21), 'msc' indicates the insertion in *mesoclisis*; the *preverbal position* is marked by 'prv' and the *postverbal position* by 'psv'. We remind that Albanian non-active clitic *u* and the dative *i* have the same distribution as the object clitics. As to mesoclisis in the Romance dialects we have considered, the phenomenon implies the co-occurrence of the accusative in enclisis, suggesting a partially different mechanism. Moreover, in these dialects mesoclisis involves also the locative form *tʃə* 'here' syncretic for the 1st plural person 'us', and the 3rd person dative *nə* 'to.him/her'.

(21)   Clitics in 2nd plural person of imperative

	1st sg	3rd acc/dat/NA	1st sg+3rd acc	3rd dat+3rd acc
Gjirokastër	msc	msc	msc	msc
Arbëresh	msc	psv	msc	psv
Shkodër	prv	msc/psv	prv	msc/psv
Terranova P.	psv	psv	msc (1st)	msc (dat/Loc)
Morano	psv	psv	msc (1st)	msc (dat/Loc)

The most immediate generalization evidenced by (22) concerns the reciprocal distribution of the clitics in imperatives, whereby we have the following ordering, in (22):

(22)   Deictic clitics/dative/locative – 3rd Person and NA
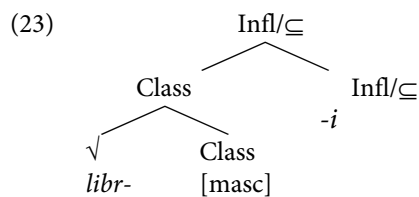
The distributional variation in (3–7) and in (13–14), contrasting 1st person objects with deictic content, and 3rd person clitics, can be traced back to DOM Differential object marking phenomenon (Comrie 1979, Croft 1988, Bossong 1991). In the typological literature, the essence of DOM is that certain types of objects, of which participants in the discourse are the
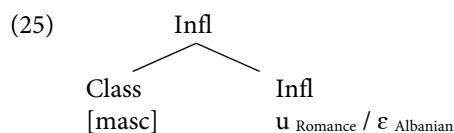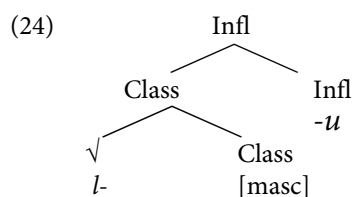
fundamental subset, are overtly marked (Aissen 2003, Bárány and Kalin in press). This phenomenon is treated in terms of referential properties, essentially animacy, definiteness and specificity, topicality expressed by means of a hierarchy regulating the distribution of grammatical functions in case systems (Dixon 1994, Kiparsky 2008, Aissen 2003, Bárány and Kalin in press), whereby the prominent elements in the scale are favoured in assuming the overt morphological mark. Depending on such basic factors, the prominence of 1st and 2nd person pronouns is easily derived. Specifically, the split between 1st and 3rd clitics seems to reflect their different interpretive status in relation to the syntactic representation of the pragmatic content (Speech Act in Speas and Tenny 2003). More precisely, deictic pronouns are interpreted independently of the event they are participants to, as anchoring to the discourse universe. 3rd person elements (on a par with nouns) are anchored to the event introduced by the verb (Manzini and Savoia 2005, 2011), in other words, they are interpreted in relation with the argumental structure of the scene/action lexicalized by the verb.

In Albanian as well as in South Italian varieties ones, this difference is expressed by the fact that 1st and 2nd person OCls do not distinguish the accusative from the dative/oblique, but show the oblique form also in transitive contexts. Contrary, in Romance varieties, 3rd person clitics separate accusative from oblique forms, while non-clitic 3rd person forms lack any case morphology, see the pronouns *jiḍḍu*/*iḍḍa* 'he/her' in Morano dialect. More precisely, 1st person clitics are inserted in the string independently of their argumental role in the event (v, VP), and have an oblique form that, following Manzini *et al.* (in press), is required in order to be interpreted in the sentence. The idea is that the oblique is not selected by the verb but it is inherent to the DOM element, a sort of 'possessor' of the event. In any case, the oblique complies with the autonomous interpretation of DOM elements and their occurrence in the modal domain of the verbal element. It is interesting to note that in Romance dialects in section 4, mesoclisis brings together 1st person clitics and locative *tʃə* and dative *nə*. The deictic nature of *tʃə* can account for its connection with the other referential clitics; as to the dative *nə*, we must conclude that it is treated like 1st person clitics, inherently oblique.

As a first step, we will consider the morphological structure of pronouns. In keeping with the model proposed in Manzini and Savoia (2018), Savoia *et al.* (2017, 2018), specifically concerning nouns, inflectional structures are built in the syntax. Thus, inflected nouns are analysed as the result of a syntactic Merge operation that combines a lexical root with gender (feminine/masculine), other classificatory properties and number. Labels are mostly self-explanatory, such as '√' for the category-less root (Marantz 1997) with predicative content (Higginbotham 1985), and 'N Class' to host gender and eventually number specifications. 'Infl' is the label for the vocalic morpheme which in romance varieties externalizes gender and/or number in terms of inflectional class. So, for instance, *libr-* 'book' in Italian combines with nominal class specifications including gender (e.g. masculine, feminine), which restricts the argument of the predicative base. A vocalic morpheme encodes properties that include both nominal class and declension class, for instance *-o* in *libr-o* 'book'. As for the plural, we assume that it corresponds to the part-whole/inclusion property, $[\subseteq]$, whereby the denotation of the root can be partitioned into subsets of individuals, as in (23). In other words, plurality denotes a subset (Chierchia 1998).

(23)
```
                    Infl/⊆
              ┌───────────┴───────┐
           Class                Infl/⊆
        ┌────┴────┐               -i
        √       Class
      libr-     [masc]
```

The case of clitics is interesting in the sense that their structure can be analyzed in the same way as a full noun, with the lexical base, *l-* in the enclitic forms of Romance dialects, combined with Class and inflection, as in (24). In Albanian and generally in proclitic forms of Southern Italian varieties (see (13e) and (14e)), clitics have a simpler structure lacking the root, i.e. coinciding with the class and inflection elements, as in (25). Therefore, Romance dialects alternate two different lexical entries for clitics, according to the sentence structure, i.e. between enclitic vs. proclitic occurrence, and in the imperative between positive vs. negative form.

(24)
```
                   Infl
             ┌──────┴──────┐
          Class          Infl
        ┌───┴───┐         -u
        √     Class
        l-    [masc]
```

(25)
```
              Infl
         ┌─────┴─────┐
       Class       Infl
      [masc]    u Romance / ε Albanian
```

As noted, the 1st/2nd person clitics occur both in transitive and intransitive contexts without distinguishing accusative and oblique. We conclude that they realize the oblique element, i.e. DOM lexicalization of deictic IA. It is no accident that in the clitic clusters and with respect to the verb, including imperative form, they have the same distribution of dative and locative elements. We obtain the representation in (26)

(26)
```
            Infl/⊆
         ┌────┴────┐
         √       Infl
         m      (i/ə)⊆
```
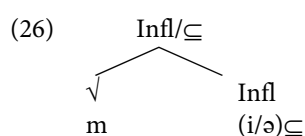
In (26) the part-whole relation characterizes oblique as well, by assuming that the inclusion relation between two arguments is the basic relation underlying prepositions and oblique case (Manzini and Savoia 2011, Franco and Manzini 2017, Savoia *et al.* 2019). Also in the case of verbs, they can be understood as the result of the syntactic Merge operation that combines a lexical root with voice, modal-aspectual and agreement inflections. Labels are mostly self-explanatory, such as '√' for the category-less root and 'Infl' for the morphemes that externalize nominal (person and number) features of the verb. We represent the internal structure of the 2nd plural of imperative, for instance *ʃixni* 'see.pl' (Gjirokastër) in (27), where 'Mood' corresponds the imperative content.

(27)

```
              Infl
            /      \
        Mood        Infl
       /    \       ni_x [2^{nd} pl]
      √       Mood
    ʃix_{x,y}  Imperative
```
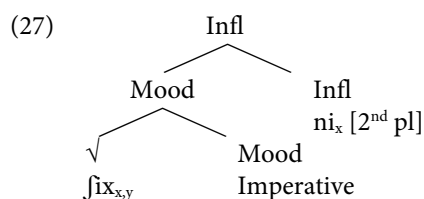
(27) combines the verbal root *ʃix* 'see', selecting two arguments, EA *x* and IA *y*, the counterfactual property of order (Imperative) and the person Inflection *ni*, on its own saturating the EA, as in (27). The proposal to introduce a word-internal slot for the modal content is substantiated by the fact that there are languages where imperative brings along specialized types of inflection. An example at hand is provided by the variety of Shkodër, where the 2nd person imperative of verbal roots ending in a consonant, has a short vocalic stressed nucleus differently from the otherwise coincident form of the reduced participle (Manzini and Savoia 2007), as in the comparison illustrated in (28a) for imperative vs (28b) for reduced participle.

(28)   a.   m   vɛʃ
            me   dress
            'dress me!'
       b.   jam   ve:ʃ
            I.am   dressed / I dressed myself'
            'I am dressed'

<div align="right">Shkodër</div>

Morpho-phonological differences of this kind support the idea that mood or other aspectual/modal verbal category can be registered by formal means. In this case, the vowel duration (and its aperture degree) is involved.

A point discussed on several occasions and articles (recently in Baldi and Savoia 2019) concerns the framework we keep to, in which all lexical material, including inflectional material, is associated to interpretable contents; this proposal is not substantially different from the conception of Agree in Chomsky (2001) insofar as it expresses the identity between features under locality (Minimal Search). As a consequence, there is no Agree rule triggered by the need for a probe to interpret/value its features and, more basically, our model excludes uninterpretable features and probe-goal induced movement, i.e. the fundamental mechanisms of cartographic explanations (see Chomsky *et al.* 2019). Agreement works by lexicalizing phi-feature bundles identifying the same argument, i.e., ultimately, denoting a single referent (Manzini and Savoia 2005, 2018).

6. Imperative, clitics and negation

The distribution of the OCls in imperatives highlights the relation between inherent interpretive properties of personal pronouns and the Phases, i.e. the licensing domains of the clause. Chomsky (2001) identifies phases with lexical subarrays, i.e. structures, computed at the SM and C-I interfaces as the result of the operation of Transfer. The procedure is constrained by the Phase Impenetrability Condition (PIC) in (29). The idea is that in a structure [$_{ZP}$ Z…[$_{HP}$

α [H YP]]], where Z and H are heads, the complement YP of H is not accessible to operations at ZP (Richards 2011).

(29)   *PIC*

The domain of H is not accessible to operations at ZP; only H and its *edge* are accessible to such operations
Chomsky (2001: 14)

Chomsky (2007, 2013, 2020) assumes the existence of two phases, CP and vP. The CP phase implies inheritance of features from the phase head C to the lower head T. Furthermore "the inheritance mechanism is simplified if it is generalized to phase heads generally, not restricted to C but extended to v* as well […] Therefore V (or R) must receive φ-features from v*. It follows that just as a nominal phrase can raise to SPEC-T within CP, so it should be able to raise to SPEC-V within v*P" (Chomsky 2007: 20–21). The distribution of Albanian clitics with respect to the Phases CP and vP may be depicted in the schema in (30), where the lexical verb and the 1st person clitic (PCl) belong to the same phase, while the 3rd person clitic (OCl) is associated to the v domain.

(30)   a.   Mood/CP phase:

C	PCl	/	PCl	C	T
verb root	mə	/	mə	verb root	Inflection

b.   vP phase:

OCl	v
ɛ / i / u / m-a	

We are suggesting that imperatives in C externalize the modal operator (imperative force) by taking scope over all the lexical material. Assuming a suggestion in Roberts (2010) as regards the object clitics in Romance languages[3], we may surmise that OCls are the true head of agreement for *v* phase. If we extend this intuition to Albanian inasmuch it is endowed with OCls, the OCl in mesoclisis is to be viewed as the agreement head for vP phase. Furthermore, in imperatives it preserves a position overtly associated to the vP domain. As to the hypothesis that the inflectional morpheme *-ni* may lexicalize *v*, we note that *v* is accessible to the operations at CP and it substantially provides the morphological instantiation of phi-features licensed by T.

Note that in Albanian varieties the clusters *1st/2nd/dative+accusative* select the *a* form of the accusative, as in *jɛp-m-a-ni* 'give (2pl) it to me' for Gjirokastër in (3d'). In isolation the accusative form for the 3rd singular is generally *ɛ*, whereas *a* can be identified as a sandhi form of the 3rd singular person accusative clitic. This suggests that in imperatives the sequences *m+a* 'me-it' and *i-a* 'him/her+it' are real clusters, confirming that no Phase boundary is involved between *m/i* and *a*. In other words, we are induced to assume that clusters are however licensed in the C-T domain.

---

3   Roberts (2010: 57, with adaptations) deals with OCls as bundles of phi-features on the edge of the v phase, as for instance in (i), cf. also Mavrogiorgos (2006).

(i)   a.   …*le voit*

'he sees him/it'

b.   [v* *le*[iφ] [v* *voit* V [v* V, uV, uφ]]]]

6.1. DOM and cliticization in Albanian

Let us consider the example in (4d') for Firmo, *ɔ-m-ni-ɛ* 'give it to me' instantiating both the DOM effect and mesoclisis. The possibility that lexical material could be inserted between root and inflection has been treated in the DM framework by Halle and Marantz (1994), Harris and Halle (2005), Arregi and Nevins (2018), in terms of rules manipulating the morpho-syntactic features associated to the terminal nodes of syntactic structure before inserting the lexical items. Our analysis diverges from this model on essential points, by assuming that lexical items (including inflections) are endowed with interpretable content on the basis of which they are inserted in the syntactic structure; in other words, as suggested, inflectional structures are built in syntax. This approach allows us to treat the phenomenon of mesoclisis between root and inflection, as in (31), in a natural way. Indeed, we can expect that clitics and inflectional morphemes can interact in order to lexicalize relevant interpretive properties.

(31)   Imperative Force

$$
\begin{array}{l}
\sqrt{}\\
\text{ɔ}\quad 1^{st}\text{Cl}\qquad\text{TP}\\
\qquad m\\
\qquad\quad\text{T}\qquad\text{vP}\\
\qquad\quad -ni\\
\qquad\qquad\text{v}\qquad\text{VP}\\
\qquad\qquad\qquad 3^{rd}\text{Cl}\\
\qquad\qquad\qquad\quad\varepsilon
\end{array}
$$
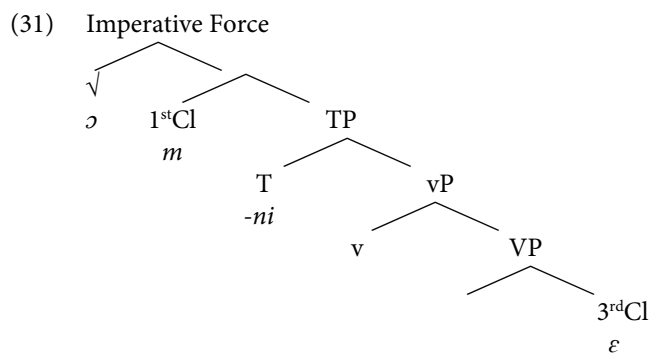
Firmo

In (31), all elements except the verbal root are able to be referred to the real world. What we see is that deictic clitics, as *m* 'me', may occur freely in the immediate context of the verbal root, deploying its capability to be interpreted independently of the verb/event. In the varieties in (3) and (4–5), it occurs in a position immediately to the right of the root, while in the Shkodër variety it occurs before the root. In any case, it seems to be associated to the phase of the root, conventionally identifiable with the domain of C. In (31) we label this domain *Imperative force*. We have tied this distribution with DOM, as far as 1st person clitics are inserted in the string independently of their argumental role in the event (v, VP).

The data from San Costantino in (5c') show that the mesoclisis of *m* entails the lexicalization of the specialized inflection of 2nd plural person *-ni*. In this dialect the 2nd plural person of imperative in isolation has typically the inflection *-i*, coinciding with the inflection of the indicative present, as in (32a,b) vs (32b'), where *-ni* is inserted.

(32)   a.   ju   pic-i   miʃ-t   *indicative*
            you   cook-2PL   meat-the
            'you cook the meat'
        b.   pic-i   miʃ-t   *imperative*
            cook-2PL   meat-the
            'cook(2pl) the meat'
        b'.   pic-   mə-   n-i-   ɛ
            cook-   me   2PL   it
            'cook(2pl) it for me'   San Costantino

As to *-i,* it is a syncretic form occurring in different person contexts, for example in the 1st plural of the present indicative, e.g. *mbuʎɔ-m-i* 'we cover', in the 3rd person of the past, *mbuʎɔ-i* '(s)he covered'. On the contrary, *-n-i* is specialized for the 2nd plural person reference. We could conclude that in imperatives its insertion is favored in order to externalize the Phase boundary between CP and vP, as in (33) (see section 6.2).

(33)   Imperative Force
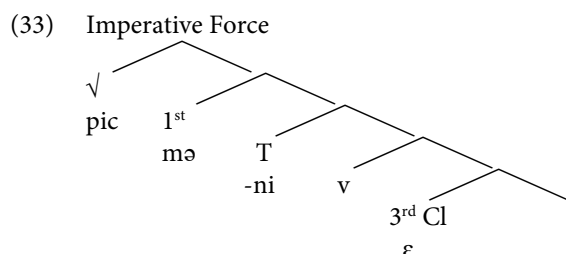
San Costantino

The 3rd person element has the accusative morphology in the domain of the event (v/V), as shown by Arbëresh varieties where the object clitic follows *-ni.* Differently, in Shkodër variety, it precedes *-ni* in the internal context and in Gjirokastër variety it is in turn placed in mesoclisis. It is reasonable to relate the distribution of the 3rd person clitic with the fact that it needs to be licensed by v, as a participant to the event, as usually. More precisely, in the presence of a non-veridical operator such as imperative mood, definiteness is lexicalized by licensing the pronoun out of the immediate scope of the operator. This interpretive effect is reflected in positioning the 3rd person clitic in the licensing domain of v, in enclisis or in mesoclisis. If that is the case, we conclude that the inflectional element *-ni* is available also to externalize v. This seems confirmed by the reduplicative structures of Shkodër in (7d), e.g. *nɛp-ni-j-a-(ni)* 'give (2pl) it to him/her', suggesting that *-ni* can occur both as the inflectional part of the verb (T) and the lexicalization of v, as in (34).

(34)   Imperative Force
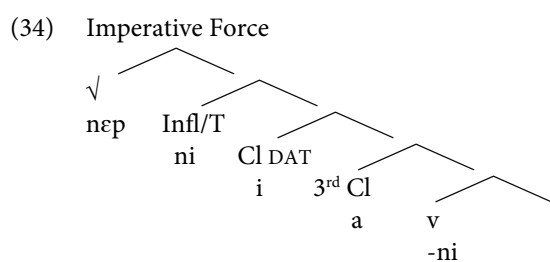
Gjirokastër

We suggest that in (34) the inflection of the verb, inherently referential (2nd person), supplies the deficient T in identifying the person and number properties of the external argument.

In the dialect of Gjirokastër, the cluster *1st person + 3rd person*, as in *sil-m-a-ni* 'send (2pl) it to me' in (3d'), is placed between the root and the inflection. We conclude that the deictic content of the 1st person clitic and the 2nd plural person *-ni* inflection are able to provide the interpretability conditions for the 3rd person clitic, as in (35).

(35)   Imperative Force
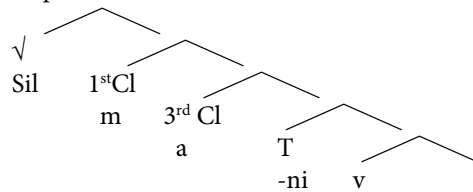
Gjirokastër

Substantially the same treatment can be extended to the data of Shkodër in (7a',b',d'), where the cluster *1st person + 3rd person* is legitimized in the pre-verbal position, as in *m a nɛp-ni* 'give (2pl) it to me' in (7d'), as in the structure in (36). Again, the 1st person clitic has referential properties sufficient to make the 3rd person clitic interpretable in the modal domain.

(36)   Imperative Force
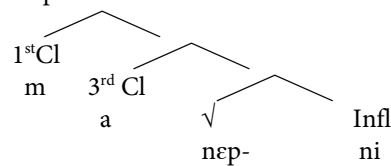
Shkodër

Synthesizing, the surface clitic order externalized in imperative contexts depends on whether the clitic is licensed by the universe of discourse (1st person) or by the event (v-3rd person object).

6.2. Romance varieties

Let us consider now the Romance systems presented in sections 4 and 4.1, where mesoclisis of 1st person and dative/locative clitics is selected only on condition that one 3rd person clitic is present in the final position, as in (37) for Morano (from (14c)).

(37)   Imperative Force
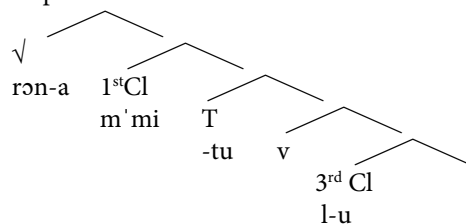
Morano

This distribution leads to the conclusion that deictic clitics, i.e. 1st person, locative and dative, admit both mesoclisis and enclisis by virtue of their referential content, which makes it possible to interpret them in different domains. On the contrary 3rd person clitics occur in the immediate domain of v. We can explain this restriction by assuming that in these dialects the agreement head of *v* is satisfied by elements endowed with referential content, i.e. deictic clitics, the 1st person element *mi*, or 3rd person clitics *l-u / l-a / l-i* endowed with the definiteness base *l-* (Manzini and Savoia 2017). Also in this case, a complete referential property of OCls is required in imperative contexts so that mesoclisis is admitted only if this requirement is fulfilled. This

seems to hold true also for other types of doubling, for instance the ones in (9) for San Marzano and in (19) for San Severo, illustrated in (38).

In the sentences in (19), reduplication of the object clitics, as *di-mma-ˈm-illə* 'tell (2sg) it to me' in (19a') is confined to contexts of 2ⁿᵈ singular person, where the lexicalization of the argumental properties (inflection in T) is missing, as in (38).

(38)   Imperative Force
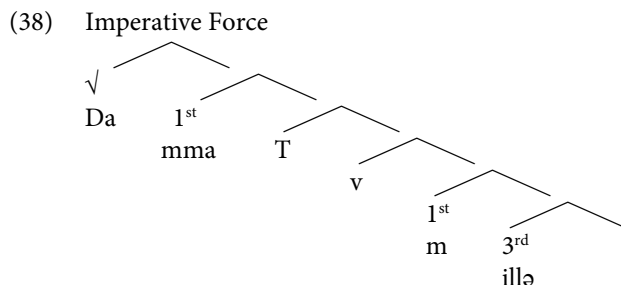
                                                                San Severo

The sentence is into the scope of the modal form (the verbal root). We find the first occurrence of the 1ˢᵗ person clitic in the immediate domain of the imperative operator and the sequence *ˈillə* 'me-it' as the specialized externalization of the participants to the event in the v Phase.

6.3. *Negative contexts*

In negative contexts, OCls are pre-verbal both in Albanian and Romance varieties. Negation can be treated as an operator "introducing a quantification over the internal argument" of the elementary event VP (Manzini and Savoia 2017: 92). This conclusion is supported by numerous phenomena in which negation and internal argument syntactically interact. Hence, the different position of clitics can mean that in the presence of negation the verb does not lexicalize the imperative force, externalized by other means, whereby the pre-verbal order valid in non-pragmatic domains is applied as suggested in (39). In San Marzano dialect, object clitics are doubled in proclisis and in enclisis in negative contexts, as shown by the example *mɔsə m ɛ hua-nni-ɛ* 'do not tell it to me' in (9b).

(39)   Imperative Force
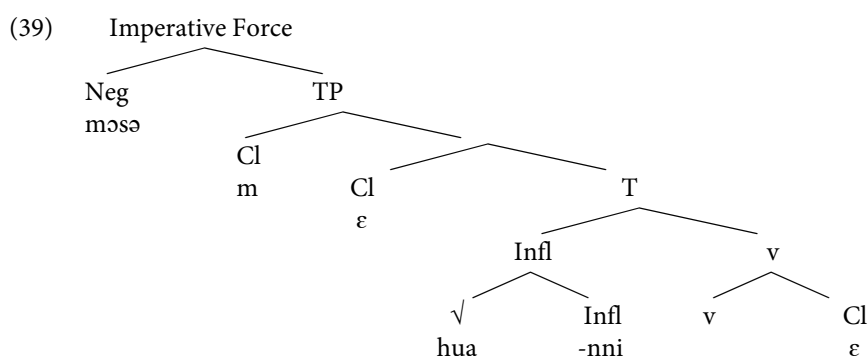
                                                                San Marzano

Negation and modal contexts (imperatives, interrogatives) are core instances of what Giannakidou (1998, 2011) calls non-veridical contexts "veridicality is a property of sentence embedding functions: such a function *F* is veridical if *Fp* entails or presupposes the truth of *p*. If inference to the truth of *p* under *F* is not possible, *F* is nonveridical" (Giannakidou 2011:

1674). We have seen in Section 3 that in many languages commands addressed to the speaker or to 3ʳᵈ person are lexicalized by hortatives/subjunctives by exploiting the event variable introduced by these forms, excluding a veridical reading. In other words, both subjunctive and imperative express a relation P(x,y) where the eventive properties of P are indefinite, not relying on truth conditions. As to negation of imperatives, typological studies document different results. Many languages allow imperative to be negated like other types of clauses (Aikhenvald 2010). This holds true for different linguistic groups, and, among Romance varieties, for example in French, in a sub-set of Italian dialects and in some Romansh varieties, as in (40).

(40)   klɔma ɛl
         call   him
         'call him'

         bɪtʃa klɔma ɛl
         neg   call   him
         'do not call him'

<div align="right">Donat (Grisons)</div>

As Han (1999) stresses, negation is never about order but about the propositional content, whereby the interpretation ¬ *Imperative Force (p)* is excluded. This explains why negation is incompatible with imperative forms in many languages (Alcázar and Saltarelli 2014), although in others, such as Albanian, the combination is admitted (Aikhenvald 2010). Hence, in many languages the direct negation of the imperative form is avoided and substituted by the insertion of verbal forms devoid of referential properties, available for non-veridical interpretation, such as infinitive. In Standard Italian and generally in Central and Southern Italian dialects the 2ⁿᵈ singular person of negative imperative is realized by a form devoid of tense and phi-features. i.e. infinitive in (17a) for Terranova, or gerund, in (20b') for San Severo (see the survey in Manzini and Savoia 2005). In cartographic descriptions, the insertion of a form devoid of agreement features is explained as a sort of suppletion to the positive imperative of 2ⁿᵈ person. Zanuttini (1997) explains this phenomenon as due to the nature of the basic form of the 2ⁿᵈ singular person of imperative, understood as unable to check the mood category in presence of the negation. Also Rivero (1994) and Giannakidou (1998) explain the incompatibility of negations and imperatives assuming that the presence of the negative head blocks the movement of the verb to the Mood higher position preventing it from licensing mood.

What the data suggest is that in Romance varieties negative operator requires an indefinite lexicalization of the event, excluding phi-features specifying the EPP argument and the referential coordinates of the event. However, what we see is that in Romance dialects negative imperatives are expressed by the infinitive (or gerund), that is an indefinite non-veridical form, excluding referential phi-features and eventive coordinates; 2ⁿᵈ and 1ˢᵗ plural persons coincide with the indicative ones. In Albanian, the negative marker specialized for modal contexts is inserted. Finally, in all the varieties we examine here, the pre-verbal position of clitics is attested, which we have connected to the fact that in negative contexts the verb does not lexicalize the pragmatic force and the pre-verbal order is restored. In (16)-(17) we have seen that in Calabro-Lucanian dialects the *l-* forms of clitics are inserted in negative contexts, as in (41); *l-* accusative clitics are generally selected in negative domains, including declarative sentences (cf. Baldi and

Savoia 2019). The definiteness root *l-* provides a complete referential content in contexts where the pronoun is out of the scope of negative operator, as in the imperative contexts (see Manzini and Savoia 2017).
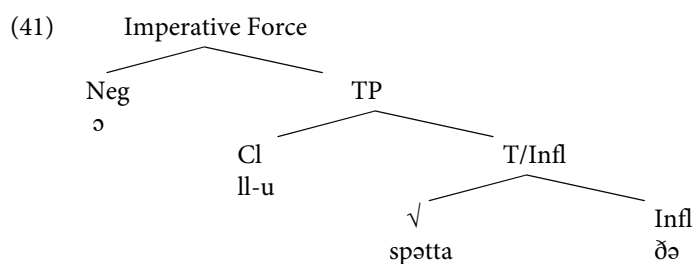
(41)   Imperative Force
         Neg   TP
          ɔ
                    Cl   T/Infl
                   ll-u
                           √   Infl
                          spətta   ðə
                                                    Terranova

In the case of Cardeto in (21) the infinitival inflection *-ri* is extended to all persons of imperative being added to the 1st plural or 2nd plural inflection, as in (42) for *nɔn tʃ-u da-mu-ri* 'do not give (1pl) it to her/him' (Cardeto).
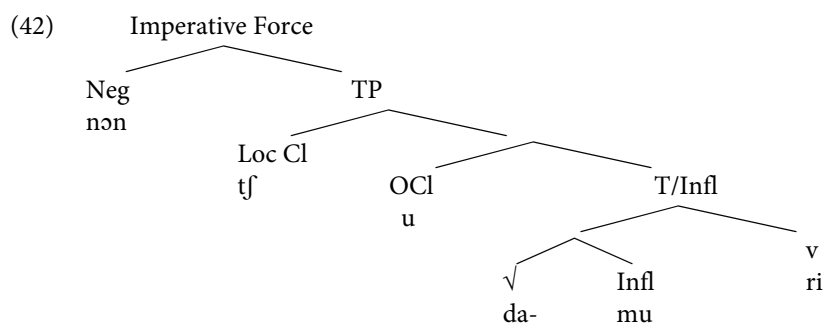
(42)   Imperative Force
         Neg   TP
         nɔn
                  Loc Cl
                    tʃ   OCl   T/Infl
                                 u
                                         √   Infl   v
                                        da-   mu   ri
                                                    Cardeto

A possible suggestion comes from the discussion concerning (37), whereby the infinitival inflection has sufficient referential properties to satisfy the requirements of *v* in contexts of the imperative quantification.

7. A brief recapitulation

Keeping the preceding discussion in mind, the distribution of OCls and the plural inflection can be connected to the externalization of the modal properties of imperative. We have assigned the initial position of the verb to the scope position corresponding to the pragmatic force and the speech act restrictions characterizing the imperative form. The non-veridical nature of the imperative sentences is highlighted by the form of the verb, that coincides with the root in the singular and, at least in a subset of contexts, also in the plural, where the root is separated from the inflection. In other words, the imperative form lexicalizes the only predicative content of the verb.

In mesoclisis structures, a type of person split emerges, whereby person clitics (1st person ones) occur close to the verbal root, suggesting that they share the same domain as the root. We connect this with the fact that they are interpretable on their own, on the basis of their deictic properties. On the contrary, 3rd person object clitics are positioned on the right of the verb and, specifically, of the 2nd plural morphology, so reflecting the necessity for 3rd person elements to

be licensed in the domain of the event (v). The properties inherent in the two series of clitics account for the distributional phenomena observed in imperatives, including the fact that in some varieties mesoclisis is triggered with the 1st person clitic and possibly with clusters. The Phase model permits to represent the distribution of verbal root, object clitics and verbal inflection in terms of principle, uncovering the different licensing domains of the two kinds of clitics and of the two parts of verbal forms, i.e. root and inflection.

References

Aikhenvald, A. Y. 2010. *Imperatives and commands*. Oxford: Oxford University Press.

Aissen, J. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21: 435–483.

Alcázar, A. and M. Saltarelli. 2014. *The syntax of imperatives*. Cambridge: Cambridge University Press.

Arregi, K. and A. Nevins. 2018. Beware Occam's syntactic razor: Morphotactic analysis and Spanish mesoclisis. *Linguistic Inquiry*, 49(4): 625–683.

Baldi, B., and L. M. Savoia. 2018. Morphosyntactic reorganization phenomena in Arbëresh dialects: The neuter. *Quaderni di Linguistica e Studi Orientali QULSO* 4: 109–130.

Baldi, B., and L. M. Savoia. 2019. Auxiliary selection in Italian dialects: Person split, OCls and raddoppiamento fonosintattico. *Quaderni di Linguistica e Studi Orientali QULSO* 5: 55–105.

Bárány, A. and L. Kalin. in press. *Introduction*. In A. Bárány and L. Kalin (eds.), *Case, agreement, and their interactions: New perspectives on differential object marking*, 1–25. Berlin: De Gruyter.

Bossong, G. 1991. Differential object marking in Romance and beyond. In D. A. Kibbee and D. Wanner (eds.), *New analyses in Romance linguistics*, 143–170. Amsterdam: John Benjamins.

Chierchia, G. 1998. Partitives, reference to kinds and semantic variation. In A. Lawson (ed.), *Proceedings of Semantics and Linguistic Theory*, vol. VII, 73–98. CLC Publications, Ithaca,

Chomsky, N. 2001. Derivation by Phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, 1–54. Cambridge, MA: The MIT Press.

Chomsky, N. 2007. Approaching UG from below. In U. Sauerland, and H.-M. Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the view from semantics*, 1–29. Berlin: Mouton de Gruyter.

Chomsky, N. 2013. Problems of projection. *Lingua* 130: 33–49.

Chomsky, N. 2020. Puzzles about phases. In L. Franco and P. Lorusso (eds.), *Linguistic variation: Structure and interpretation*, 163–167. Berlin: De Gruyter Mouton.

Chomsky, N., Á. Gallego, and D. Ott. 2019. Generative Grammar and the faculty of language: Insights, questions, and challenges. In Á. Gallego and D. Ott (eds.), *Generative syntax: Questions, crossroads, and challenges*, Special issue of *Catalan Journal of Linguistics*: 229–261.

Comrie, B. 1979. Definite and animate direct objects: A natural class. *Linguistica Silesiana* 3: 13–21.

Croft, W. 1988. Agreement vs. case marking and direct objects. In M. Barlow and C. A. Ferguson (eds.), *Agreement in natural language: Approaches, theories, descriptions*, 159–179. CSLI.

De Angelis, A. online. *Microvariazione, diacronia e interferenza: Due case-studies dall'Italia meridionale estrema.*

Dixon, R. M. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Franco, L., and M. R. Manzini. 2017. Instrumental prepositions and case: Contexts of occurrence and alternations with datives. *GLOSSA* 2: 1–37

Giannakidou, A. 1998. *Polarity sensitivity as (non)veridical dependency*. Amsterdam: John Benjamins.

Giannakidou, A. 2011. Negative and positive polarity items. In K. von Heusinger, C. Maienborn, and P. Portner (Eds.), *Semantics: An international handbook of natural language meaning*, *Volume 2,* 1660–1712. Berlin: Mouton de Gruyter.

Halle, M., and A. Marantz. 1994. Some key features of Distributed Morphology. In: A. Carnie, H. Harley and T. Bures (eds.), *Papers on Phonology and Morphology. MIT Working Papers in Linguistics* 21: 275–288.

Han, C.-H. 1999. Deontic modality, lexical aspect and the semantics of imperatives. *Linguistics in the Morning Calm* 4: 479–495.

Harris, J., and M. Halle. 2005. Unexpected plural inflections in Spanish: Reduplication and metathesis. *Linguistic Inquiry* 36: 195–222.

Higginbotham, J. 1985. On semantics. *Linguistic Inquiry* 16: 547–621.

Kiparsky, P. 2008. Universals constrain change; change results in typological generalizations. In J. Good (ed.), *Linguistic universals and language change*, 23–53. Oxford: Oxford University Press.

Kuryłowicz, J. 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter Universitätsverlag.

Ledgeway, A., N. Schifano, and G. Silvestri. 2016. The negative imperative in southern Calabria: Spirito greco, materia romanza again? Paper presented at *FARM Workshop, University of Bucharest*, 25–26.11.2016.

Loporcaro, M. 1995. Recessività sintattica dell'infinito e vitalità morfologica della desinenza infinitivale in alcuni dialetti del reggino. In R. Ajello, and S. Sani, *Scritti linguistici e filologici in onore di Tristano Bolelli*, 331–358. Pisa: Pacini.

Manzini, M. R., and L. M. Savoia. 1999. The syntax of middle-reflexive and object clitics: A case of parametrization in Arbëresh dialects. In: M. Mandalà (ed.), *Studi in onore di Luigi Marlekaj*, 283–328. Bari: Adriatica.

Manzini, M. R., and L. M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.

Manzini, M. R., and L. M. Savoia. 2007. *A unification of morphology and syntax. Investigations into Romance and Albanian dialects*. London: Routledge.

Manzini, M. R., and L. M. Savoia. 2011. Mesoclisis in the imperative: Phonology, morphology or syntax? *Lingua* 121: 1101–1120.

Manzini, M. R., and L. M. Savoia. 2017. Enclisis/Proclisis alternations in Romance: Allomorphies and (re)ordering. *Transactions of the Philological Society* 115: 98–136.

Manzini, M. R., and L. M. Savoia. 2018. *The morphosyntax of Albanian and Aromanian varieties*. Berlin/Boston: De Gruyter Mouton.

Manzini, M. R., and L. M. Savoia. Forthcoming. Person splits in Romance. Implications for parameter theory. In: A. Bárány, T. Biberauer, J. Douglas, and S. Vikner (eds.). *Syntactic architecture and its consequences II: Between syntax and morphology*. Berlin: Language Science Press.

Marantz, A. 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: 201–225.

Mauck, S., and R. Zanuttini. 2005. The subjects of English imperatives. In C. Brandstetter, and D. Rus (eds.), *Georgetown University working papers in theoretical linguistics. Volume IV*, 53–85. Washington, DC: Georgetown University Department of Linguistics.

Mavrogiorgos, M. 2006. The syntax-morphology interface is at the edge: Evidence from Greek clitics. *GLOW Newsletter* 56.

Platzack, C., and I. Rosengren. 1998. On the subject of imperatives: A minimalist account of the imperative clause. *The Journal of Comparative Germanic Linguistics* 1: 177–224.

Portner, P. 2004. The semantics of imperatives within a theory of clause types. In R. Young (ed.), *SALT* XIV, 235–252, Ithaca, NY: Cornell University.

Richards, M. D. 2011. Deriving the Edge: What's in a Phase? *Syntax* 14: 74–95.

Rivero, M.-L. 1994. Negation, imperatives and Wackernagel effects. *Rivista di Linguistica* 6: 39–66.

    (1)   Roberts, I. 2010. *Agreement and Head Movement*. Cambridge, MA: The MIT Press.

Savoia, L. M., B. Baldi, and M. R. Manzini. 2018. Sigmatic plurals in Romance varieties spoken in Italy and their interaction with -i plurals. *Linguistics Beyond and Within* 4: 141–160,

Savoia, L. M., B. Baldi, and M. R. Manzini. 2019. Asymmetries in plural agreement in DPs. In J. Emonds, M. Janebová, and L. Veselovská. (eds.), *Language use and linguistic structure proceedings of the Olomouc Linguistics Colloquium 2018*, 204–224. Olomouc: Palacký University Olomouc (ONLINE)

Savoia, L. M., M. R. Manzini, L. Franco, and B. Baldi. 2017. Nominal evaluative suffixes in Italian. In R. D'Alessandro, G. Innàccaro, D. Passino, and A. M. Thornton. (eds.), *Di tutti i colori. Studi linguistici per Maria Grossmann*, 283–300. Utrecht: Utrecht University Press.

Speas, P., and C. Tenny. 2003. Configurational properties of point of view roles. In A. DiSciullo (ed.), *Asymmetry in grammar*, 315–344. Amsterdam: John Benjamins.

Zanuttini, R. 2008. Encoding the addressee in the syntax: Evidence from English imperative subjects. *Natural Language and Linguistic Theory* 26(1): 185–218.

Fortis-lenis distinction of fricatives and plosives in Welsh – Phonetically distinctive factors

Michał Baran

University of Szczecin, Poland

Abstract

This paper continues a research project aimed at proving that the fortis-lenis distinction is more appropriate for studying the Welsh language than the voiced-voiceless divide. Previous analyses of articulatory timing revealed potentially distinctive features characteristic for Welsh plosives and fricatives. Even though the phonological features such as [spread glottis] or [voice] may be used both for distinguishing between the series of plosives and fricatives, the phonetic features responsible for the distinction are likely to differ due to the different phonetic nature of the manner of articulation and the position of fricatives and plosives (Honeybone 2005: 333–334). The aim of this paper is, therefore, to draw general conclusions from analysing the two sets of factors and find some universal features or characteristics of fortis and lenis obstruents in Welsh. Based on the previous studies, the importance of aspiration and voicing length are analysed in Welsh plosives. For fricatives, friction and voicing length are studied as potentially contrastive. The main hypotheses are that (i) phonetic voicing is not decisive in distinguishing between the two series of sounds in all places of articulation and (ii) other aspects of articulatory timing such as friction, aspiration length and hold phase duration appear to be more important in the fortis-lenis distinction and form a general pattern where the fortis sounds are generally longer than their lenis counterparts.

Keywords: fortis-lenis distinction, plosives, fricatives, articulatory timing, laryngeal contrast

1. Introduction

Consonant distinctions are key in describing sound inventories of languages. They let researchers organise the sounds into groups based on common features. These features may generally belong to one of three groups: place of articulation, manner of articulation and laryngeal features. Laryngeal features constitute a laryngeal system of a given language, which is here understood as "the totality of phonological and phonetic aspects which are responsible for the observed phonetic facts, where phonology and phonetics are kept strictly apart, and yet they form two sides of the same coin" (Cyran 2011: 50). On one side of the coin there are phonological features responsible for the phonological distinction such as [spread glottis], [voice] or [constricted] and on the other, there are phonetic indicators of the aforementioned features. Most frequently, for instance, aspiration suggests the presence of the [spread glottis]

factor and phonetic voicing indicates [voice]. This paper looks at laryngeal features used in the analysis of the Welsh plosives and fricatives. There are three main treatments of the laryngeal features in Welsh as far as the aforementioned groups of sounds are concerned. Firstly, there is an approach which divides plosives and fricatives into two series, i.e. voiceless and voiced (Ball and Williams 2001). In such an approach, it naturally follows that the Welsh language exhibits [voice] as a distinctive feature, but this is not confirmed by experimental phonetics (Asmus and Grawunder 2017, Ball 1984, Jones 1984). The second approach employs the terms fortis and lenis as a phonological distinction that may be based on other features than [voice] and as a result appears to solve the main issue of the voiced-voiceless approach (Hannahs 2013; Ball 1984; Cyran 2011). However, if fortis and lenis are used as synonyms of voiceless and voiced, i.e. when the authors present the same minimal pairs and oppositions as in the voiceless-voiced distinction, such account fails to explain certain phenomena found in the Welsh language, most importantly morpheme-initial consonant mutations. Therefore, the third approach, advocated in this paper, views the fortis-lenis divide as phonologically separate from the voiceless-voiced distinction in that it results in a different division of the phonemes rendering different groups of sounds and considering different minimal pairs vital. This perspective seems to be supported by morphophonological and phonological analyses (see for example Baran and Asmus 2019), but also by experimental phonetics (Asmus *et al.* 2019). If the phonological fortis-lenis distinction is to be seen as one side of the coin of the Welsh laryngeal system, the other side of the coin, i.e. the phonetic factors responsible for that distinction have to be experimentally verified.

The aim of this paper is to summarise what has currently been achieved in the research project aimed at identifying the phonetic correlates of the fortis-lenis distinction in Welsh. Two studies, one on plosives (Asmus *et al.* 2019) and the other on fricatives (Baran 2020), are presented and the findings are compared in order to contribute to the debate focused on whether the fortis-lenis distinction hinges on the same or comparable phonetic features for both fricatives and plosives. The influence of voicing, so important in the voiceless-voiced distinction, is compared with the influence of alternative features such as aspiration and friction for plosives and fricatives respectively in order to show that for the fortis-lenis distinction, phonetic voicing is at most of minor importance.

2. Theoretical background of the study

2.1. Overview of research of laryngeal distinctions

The aim of this section is to present some examples of studies on laryngeal system organisation in order to show that despite various approaches to this issue there seems to be an agreement regarding the fact that the voiceless-voiced distinction, despite being quite wide-spread, is not the only possibility and is not always synonymous with the fortis-lenis distinction. This paper does not exclusively follow any of the approaches below but favours an eclectic approach.

Jakobson, Fant and Halle (1951) mention two most important laryngeal distinctions, namely between tense and lax consonants (and vowels) and voiceless and voiced consonants

(and potentially vowels). They see "the superposition of a harmonic sound source upon the noise structure" as the main indicator of the voiceless-voiced distinction (Jakobson, Fant and Halle 1951: 26). According to these authors, the tense-lax distinction is said to rely upon two features namely the length of the sound interval and articulatory energy with the tense sounds exhibiting a longer sound interval and a higher level of articulatory energy resulting from a greater deformation of the vocal tract from its neutral position and longer articulation. Jakobson, Fant and Halle also acknowledge that the voiceless-voiced distinction is widespread in the world and that it may alternate with the tense-lax distinction in one of four ways, i.e. (i) coexist by being its concomitant, (ii) become redundant because the tense-lax distinction is more appropriate for a given language, (iii) be relevant for only some classes of consonants of a given language and (iv) be irrelevant for the language together with the tense-lax distinction.

The theory of laryngeal realism (Honeybone 2005) contributes to the debate on laryngeal systems by suggesting that there are languages for which (i) the feature [voice] is irrelevant and proves to be useless in descriptions of the diachrony of a language and its current state and processes, (ii) the feature [spread glottis] is more appropriate and enables researchers to explain previously puzzling processes. The phonological feature [voice] logically results in the voiceless-voiced distinction while [spread glottis] is associated with the fortis-lenis, or in Jakobson Fant and Halle's terms, tense-lax distinction. The languages employing the distinctions are also called voicing and aspirating languages respectively. Honeybone (2005) also argues that the features such as [voice] and [spread] are privative and that the unmarked series is non-specified, i.e. only the marked series is phonologically specified.[1] Such claims lead to rather meaningful changes not only in phonological analyses, but they may even be used to justify changes to phonological symbols used. From this very radical point of view, Welsh plosives and fricatives would be organised in the following way:

(1)   A. fortis /p^h,t^h,k^h/ vs lenis /p^o,d^o,g^o/
        B. fortis /f^h, s^h, θ^h/ vs lenis /f^o, h^o, θ^o/

This kind of notation stresses the fact that the series of sounds in (1) differ not in the [voice] feature, but in [spread glottis] with the fortis series being specified with it. In this paper, however, for the sake of clarity, the conventional notation is employed. It should be mentioned here that the minimal pairs in the description of the laryngeal system are not always the same as phonological minimal pairs in Welsh. In this paper, a phonological minimal pair is understood as one that is supported by the morpheme-initial consonant mutation system. The morpheme-initial consonant mutations may be defined as systematic and grammaticalised changes that affect morpheme-initial phonemes yielding those that are phonetically different (Asmus and Grawunder 2017: 22). The table below presents the paradigm of changes (for a more detailed analysis of mICM see Baran and Asmus 2019).

---

[1]   The notion of markedness is, however, not fully useful with languages such as Welsh, where sound changes form sequences such as /p/ -> /b/ -> /v/.

Table 1: *Morpheme-initial mutations in Welsh*

Radical	Lenition	Nasalisation	Spirantisation
p /p$^{(h)}$/	b /b$^{(h)}$/	mh /m̥/	ph /f/
t /t$^{(h)}$/	d /d$^{(h)}$/	nh /n̥/	th /θ/
c /k$^{(h)}$/	g/g$^{(h)}$/	ngh /ŋ̥/	ch /χ/
b /b$^{(h)}$/	f /v/	m /m, m̥/	
d /d$^{(h)}$/	dd /ð/	n /n, n̥/	
g/g$^{(h)}$/	Ø	ng /ŋ/	
m /m$^{(h)}$/	f /v/		
ll /ɬ/	l /l, l̥/		
rh /r̥/	r /r/		

While in some cases, for instance in plosives, the pairs are the same as the ones present in the laryngeal system, in fricatives they are quite different. For instance, /θ/ and /ð/ are definitely contrastive within the laryngeal system of Welsh as they are both interdental fricatives and differ only in the former being fortis and the latter lenis. However, in the morpheme-initial mutation system, the two sounds are not linked to each other as /θ/ is a spirantisation reflex of /t/ while /ð/ is a lenition reflex of /d/. Therefore, they are not considered phonologically contrastive to the same extent as minimal pairs such as /p/ and /b/.

Laryngeal realism advocated by Honeybone (2005) presupposes a direct link between certain phonetic aspects of a sound and its phonological specification. For instance, the presence of full voicing would mean that the sound is specified with [voice] and also that the language is a voicing language. As a result, the theory does not allow for much arbitrariness in linking phonetics to phonology. Cyran (2011) disagrees with such a point of view and introduces laryngeal relativism as an updated version of laryngeal realism. He claims that the aforementioned phonetic cues do not always indicate the same phonological features. Voicing, for instance, may be spontaneous, passive or active and only the last one is an indicative of the [voice] feature and the language having voiceless-voiced distinction. Such point of view would not change the way Welsh sounds are presented in (1) but would allow for some degree of arbitrariness while analysing their phonetic shape.

All the aforementioned theories are relevant to this paper as they confirm that there are various possible organisation patterns of laryngeal systems and it logically follows that a successful analysis of a given language requires applying the most appropriate laryngeal distinction. Understandably, there are also some differences between the approaches presented above, but as these differences are phonological in nature, they go beyond the scope of this paper.

Jaeger (1983) in a paper devoted to Zapotec and Jawoñ devotes some space to general considerations of the phonetic nature and application of the fortis-lenis distinction. His considerations are partly based on the paper by Jakobson, Fant and Halle, mentioned before. It is claimed that the fortis-lenis divide can be used either as a basic phonological divide for languages whose consonant series do not differ in terms of voicing or as an additional distinction, overlapping with and enriching the voiceless-voiced distinction. However, according to Jaeger, in languages such as English, where the feature fortis is used in analysis of the voiceless plosives because of their aspiration, the fortis-lenis distinction is seen to be

redundant as the sounds are sufficiently distinguished on the basis of VOT which is not seen as a function of the fortis-lenis distinction by Jaeger. Four groups of potential phonetic correlates of the fortis-lenis distinction are mentioned, i.e. (1) pulmonic factors, (2) articulation factors, (3) timing factors and (4) glottal factors with groups (3) and (4) being of particular importance for this paper (see sections 2.3. and 3.1.).

Before finding the correlates, one would also have to decide whether there is a single phonetic factor to distinguish between the phonological series or whether the distinction is multifactorial. Butcher (2004) provides an overview of the most common approaches to this issue. Firstly, he points to the theories which claim that there is one single feature that differentiates between the fortis and lenis consonants. Physical effort in the form of muscular tension is claimed to be such a factor. This approach is practised for instance by Hardcastle (1973) and Ladefoged (1989). Secondly, there is an approach in which a combination of two factors is used to differentiate between the fortis and lenis consonants. These factors are glottal aperture and articulatory timing, whereas voicing and aspiration are not relevant. Such a view can be found in the works of Malecot (1970) and Stetson (1951). Finally, there is a theory on which the present research is based that considers the difference between the fortis and lenis series of sounds a combination of four parameters, namely: peak glottal aperture, articulatory timing, voicing and aspiration (see Asmus *et al.* 2019).

2.2. Overview of research of laryngeal distinction of Welsh consonants

Ball (1984) claims that (i) the voiceless-voiced distinction reflects the state of the glottis during the closure stage, (ii) the aspirated-unaspirated distinction refers to the state of the glottis during and after the release of the closure and (iii) the fortis-lenis distinction is seen as something related to the force of articulation and cannot be verified experimentally. This approach directly links certain articulatory aspects to specific distinctions, not allowing for much arbitrariness. Ball claims that both for stops and fricatives voicing is irrelevant in the initial position and may sometimes be relevant in the final position while distinguishing between the fortis and lenis series. He also states that other factors such as (i) VOT/aspiration for plosives, (ii) friction length for fricatives and (iii) preceding vowel length and quality offer considerably more conclusive marking of the fortis and lenis series. These results seem to confirm that Welsh is an aspirating language with a fortis-lenis distinction.

Jones (1984) is also interested in the laryngeal distinctions of Welsh obstruents. He focuses on stops and fricatives and omits affricates as they occur mainly in borrowings. As far as stops are concerned, he notices that the unaspirated plosives are not constantly voiced and that VOT is decisive in discriminating between the fortis and lenis series. He is also aware of potential differences in the force of articulation, but like Ball claims that there is no experimental justification. For fricatives, the same pairs of sounds as in Ball are analysed and Jones' findings are the same as Ball's as far as decisive factors are concerned.

Jones' paper would also support the idea that the Welsh obstruent inventory features a fortis-lenis distinction. However, even though both Ball (1984) and Jones (1984) employ the fortis-lenis distinction in individual sounds, it does not lead them to analyse pairs of sounds contrastive along the fortis-lenis divide supported by the mICM. The contrasts they analyse

stem from the voiceless-voiced divide or do not always form proper minimal pairs.[2] Fricatives are the best example of such a situation. As it was mentioned before, /f/ and /v/ and /θ/ and /ð/ are homorganic and contrastive as far as the laryngeal system is concerned, whether we from the fortis-lenis or voiced-voiceless perspective. However, phonologically speaking, in Welsh, /v/ is contrastive to /m, b/ word-finally (see Asmus and Grawunder 2017).[3] Word-initially, it does not normally occur in native lexemes, unless mutated and is then contrastive to /m, b/ as well. The minimal pairs such as *y ffordd* 'road (Old English borrowing)' and *y ford* 'table (Old English borrowing, mutated)' that may be presented in order to suggest a /f/ and /v/ contrast are normally borrowings. As a result, they may establish a gateway into Welsh for such a contrast by allowing to affect other sounds along the voiceless-voiced pronunciation leading to changes in Welsh phonology.

For /θ/ and /ð/ there seem to be minimal pairs such as *gwyrth* 'miracle (Latin borrowing)' and *gwyrdd* 'green (Latin borrowing)', *moth* 'moth (English borrowing)' and *modd* 'mode (Latin borrowing)', or, if one decides to focus on the words of native origin, *bydd* 'she/he will' and *byth* 'never/ever'. However, Asmus and Grawunder (2017) show that relative length can be observed here, making it difficult to establish a clear contrast. In addition, the two sounds are not contrastive in English, so its influence on Welsh should remain limited.[4]

Analysing truly phonologically contrastive pairs of sounds offers promising perspectives and when it comes to the Welsh fricative, /s/ versus /h/ constitutes such pair, but this issue is omitted both in Jones and Ball. Even though it is not actively employed in the Welsh mICM, there are certain minimal pairs, for instance *sil* 'seed' and *hil* 'breed', that reflect such a contrast. Moreover, it is well-rooted in another Insular Celtic language, i.e. Irish. Also, Asmus and Grawunder (2018) show that it is contrastive word-finally in Welsh, for instance in *rhy* 'too' and *rhys* 'embers', *plas* 'palace' and *pla* 'plague' or *lles* 'benefit' and *lle* 'space'.

Hannahs (2013) uses the voiceless-voiced terminology but is aware of the phonetic research by Ball (1984). His descriptions prove that if one uses the fortis-lenis distinction in such a way, i.e. as a synonym for the voiceless-voiced distinction then, in essence, there is no effect of this new terminology. Using the fortis-lenis distinction appears to be decisive and influential only if it leads to establishing different minimal pairs than the ones based on the voiceless-voiced distinction.

Asmus and Grawunder (2017) revealed that there is a clear correlation between vowel and coda length in Welsh monosyllables along the fortis-lenis divide.[5] The fortis-lenis divide is there understood in this radical sense that results in having different minimal pairs. The aforementioned minimal pair /s/ vs /h/, omitted in the voiceless-voiced or the quasi fortis-lenis

---

[2]   For instance, the minimal pair of *hoff* 'dear' and *hof* 'hoe' is not a proper one because the latter lexeme is 1) an English borrowing and 2) rarely used. There are also some minimal pairs which present consonants in clusters, for instance *corf* 'pommel' and *corff* 'mass'. Since this paper is focused on consonants found in simplex codas and onsets of monosyllables, such examples go beyond its scope.

[3]   An example of such a minimal pair is *tref* 'town' and *trem* 'sight'.

[4]   I would like to thank dr hab. Sabine Asmus, prof. US (Szczecin, Leipzig) for having pointed this issue out to me as well as advised on the /s/ and /h/ distinction and provided me with minimal pairs from unpublished material.

[5]   This kind of vowel-coda dependence has been studied by many researchers, for instance Chen (1970), Hogan and Rozsypal (1980) or Lisker (1986).

distinction, proves to be as decisive as /p/ vs /b/ when it comes to vowel and coda length. Contrary to common assumption, it is the consonant length which determines that of the preceding vowel, thus identifying Welsh as a consonant-driven language.

2.3. *Theoretical reasons for parameter selection*

This section introduces the parameters that were measured in this study. As aforementioned, the approach here is that the fortis-lenis distinction hinges upon four features: peak glottal aperture, articulatory timing, voicing and aspiration. All of them, apart from peak glottal aperture, were taken into consideration. Articulatory timing was linked with friction length for fricatives and hold phase duration and aspiration length for plosives. Voicing was measured both for fricatives and plosives. Understandably, aspiration was only measured for plosives. The aforementioned features are briefly defined in the following paragraphs and the reasons for their selections are explained.

Phonetic voicing is the first of the features under review and it was measured for both fricatives and plosives. Trask (1996: 381) defines it as vocal folds vibration happening while articulating a given sound and claims that it may be the one parameter that distinguishes between voiced and voiceless sounds. Even though Ball (1984) and Asmus and Grawunder (2017) suggest that the influence of phonetic voicing in Welsh is incidental at most, it should still be included in any study focused on laryngeal systems.

Aspiration, the second feature under review, is defined as "a period of voiceless breathing", that follows an obstruent (Trask 1996: 36; Stevens 1998: 451). For experimental phonetics, aspiration is friction following the release of a plosive. It is connected with the voice-onset time (abbreviated to VOT), i.e. the time between the release of the plosive to the beginning of modal voicing signifying a vowel. Since it is assumed here that Welsh is an aspirating language, this feature definitely has to be studied. Diachronically speaking, aspiration, together with general articulation length, is seen to have been decisive for the distinction between the radical fortis plosives, lenited fortis plosives and radical lenis plosives observable in Old and Middle Welsh before the lenited fortis plosives blended with the radical lenis ones forming the system known from Modern Welsh (Sybren van Sluis 2019: 75).

The third factor under review, relevant only for plosives, is the hold phase duration, also called approach. In citation forms or after a pause, /p, t, k/ are said to feature a longer hold phase than /b, d, g/, but the differences are not evident in connected speech (Ogden 2009: 98). Hold phase duration is here seen as one of the articulatory timing features and as such seems to be worth measuring because of the apparent vowel-coda duration dependence discovered by Asmus and Grawunder (2017).

The fourth factor is friction length which may be seen as an equivalent of aspiration for fricatives. Studying the friction to voice ratio offers similar potential results as the VOT. Measuring friction length also contributes to an in-depth study of the articulatory timing which as claimed in the previous paragraph, seems to be quite relevant for the distinction between fortis-lenis Welsh obstruents.

3. The study

3.1. Methodology and the participants

Both of the studies to be discussed here were conducted on the same set of recordings. The first study focused on plosives (Asmus *et al.* 2019) and the second one dealt with fricatives (Baran 2020). 31 native speakers, male and female aged 19–71, mostly from North Wales[6], who use their language at home and at work, were interviewed between 2013 and 2018. The average age of participants was 58 and the majority of them were female. As far as education is concerned, they were mostly educated standard speakers. The analysis was undertaken both in the onset and coda of monosyllabic native lexemes currently in use. The lexemes were taken from previous corpora (Asmus and Grawunder 2017) but were further amended by adding more lexemes to cover more sounds. The tokens were then placed in the carrier phrase *Dw i heb ddweud (1) ond (2)!* 'I didn't say (1) but (2)!' and were ordered in a way that every lexeme appears in the recording twice in a strong and twice in a weak prosodic position. Selected native lexemes were also recorded in normal sentences. Then, the target items were analysed with the help of the Praat software (Boersma and Weenink 2019).

In the analysis of plosives, the holding phase duration[7] (potential voicing) and aspiration were measured. It was assumed that aspiration is a period of friction extending from the release of a plosive to the onset of modal voicing. Knowing that friction following an initial lenis plosive may be partially voiced, both parameters were measured in order to establish whether their relative durations distinguish fortis plosives from their lenis counterparts. The collected data were subjected to a statistical analysis by means of a mixed-design ANOVA, which takes into consideration the influence of random effects, which include speakers and the phonological contexts.

As far as fricatives are concerned, friction and voicing length were measured. In word-initial sounds, friction length was measured from the point where the voicing of the preceding sound /d/ ceased to the beginning of modal voicing. Voicing length was measured from the first occurrence within the friction phase to the end of the friction phase. In word-final position, friction length was measured from the point where the modal voicing ceased to the point where the friction phase finished or where the following sound has clearly begun. The collected temporal data were subjected to the same type of statistical analysis as plosives.

3.2. Results

The analysis of aspiration and voicing length was conducted separately in the word-initial and word-final position. In word-initial position aspiration and voicing length were measured,

---

[6] There were only two South Welsh speakers which made it impossible to draw any conclusions regarding dialectal differences.

[7] Due to the phonetic nature of the carrier phrase the holding phase duration could only be measured accurately in the word-final position.

while in word-final position, hold phase was added to the two aforementioned features. The table below presents the results of the analysis conducted in word-initial position.[8]

Table 2: *Aspiration and voicing length in the word-initial position*

	ASPIRATION		VOICING	
	Average	Standard deviation	Average	Standard Deviation
/p/ (1)	117.6	24.5	13.6	2.4
/p/ (2)	90.3	16.6	14.2	2.2
/t/ (1)	129.3	34.6	12.2	2.1
/t/ (2)	113.4	20.7	16.9	3.3
/k/ (1)	148.2	41.4	13.5	1.9
/k/ (2)	138.6	25.8	13.8	3.1
/b/ (1)	19.9	3.1	11.3	2.8
/b/ (2)	17.3	3.1	9.5	2.5
/d/ (1)	22.8	4.2	12.1	2.5
/d/ (2)	21.8	4.6	13.7	3.1
/g/ (1)	33.4	3.7	13.8	3.6
/g/ (2)	33.4	5.1	17.1	4.4

It appears that the two series of sounds differ strongly in terms of aspiration length. The difference is visible while comparing homorganic pairs of sounds but also while looking at the fortis and lenis series as groups. As far as voicing is concerned, voicing length is not always conclusive and the differences both between homorganic pairs of sounds and the fortis and lenis series are at best slight.

Table 3: *Hold phase, aspiration and voicing length in the word-final position*

	HOLD PHASE		ASPIRATION		VOICING	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
/p/ (1)	208.2	39.8	119.6	27.7	17.4	3.1
/p/ (2)	157.3	20.2	82.4	19.4	24.6	5.9
/t/ (1)	203.6	46.3	124.7	39.4	13.4	2.4
/t/ (2)	177.8	26.4	96.9	22.5	19.2	6.8
/k/ (1)	226.9	49.5	142.6	40.8	16.1	2.6
/k/ (2)	195.7	34.4	120.1	28.8	17.3	2.7
/b/ (1)	96.2	28.9	66.8	25.4	21.2	5.3
/b/ (2)	74.6	13.7	47.3	18.6	22.7	7.8
/d/ (1)	109.5	38.7	89.6	19.7	20.4	4.6
/d/ (2)	83.4	14.2	56.6	17.1	29.4	6.3
/g/ (1)	123.3	42.2	95.9	22.3	22.1	3.9
/g/ (2)	97.3	15.6	66.2	17.9	40.1	7.2

The analysis of the word-final plosives reveals a similar pattern to the word-initial position, where voicing is not decisive in distinguishing between the two series of sounds under review.

---

[8] In this and the following tables (1) means the first position in the carrier phrase and (2) means the second position in the carrier phrase.

The differences in voicing vary from situations like /p/ (2) vs /b/ (2) where it is the fortis sound that has more voicing on average, to significantly longer voicing period in the lenis sound as in /g/ (2) vs /k/ (2). The other two parameters, i.e. aspiration and hold phase, appear to be more reliable with significantly longer periods of aspiration and longer hold phases in articulation of the fortis sounds.

In brief, analysing plosives in word-initial and word-final position suggests that voicing is not enough to distinguish between the fortis and lenis sounds in the Welsh language. Alternative measurements such as aspiration length and hold phase duration exhibit a clearer and more reliable pattern.

The analysis of fricatives was also conducted word-initially and word-finally, but regardless of the position the same two parameters were measured, i.e. friction and voicing length. However, because of distributional restrictions the sounds under review were not the same in the two positions under review. The table below presents the results of the word-initial analysis of /s, f, h, v/.

Table 4: *Friction and voicing length in the word-initial position*

	FRICTION		VOICING	
	Average	Standard deviation	Average	Standard deviation
/s/ (1)	198.8	50.3	10.5	4.6
/s/ (2)	207.6	63.2	11.4	6.5
/f/ (1)	179.2	56.4	10.9	13.3
/f/ (2)	195.8	65.6	13.7	8.2
/h/ (1)	140.4	52.6	37.2	43.7
/h/ (2)	148.1	69.1	38.4	46.7
/v/ (1)	123.2	55.8	76.4	76.6
/v/ (2)	119.4	48.5	90.5	71.5

The two series of sounds appear to significantly differ in terms of both friction and voicing length. It is worth to note that the homorganic pair /f/ vs /v/ established on the basis of the laryngeal system only exhibits the same pattern as the phonologically motivated /s/ vs /h/. The voiceless-voiced distinction that views both /s/ and /h/ as voiceless sounds would not be able to account for the observed differences in articulatory timing.

Table 5: *Friction and voicing length in the word-final position*

	FRICTION		VOICING	
	Average	Standard deviation	Average	Standard deviation
/f/ (1)	262.2	79.1	35.0	15.1
/f/ (2)	246.1	100.6	37.1	25.6
/θ/ (1)	262.0	76.8	37.1	16.9
/θ/ (2)	231.7	89.2	52.6	23.0
/v/ (1)	166.2	43.2	70.8	29.7
/v/ (2)	163.1	79.5	57.8	41.7
/ð/ (1)	156.3	56.5	52.4	18.3
/ð/ (2)	203.8	74.3	57.4	35.8

In the word-final position the analysis focused on /f, θ, v, ð/. These sounds constitute two homorganic pairs, which are not supported by the mICM system, but the pattern is still visible. The fortis sounds have considerably longer friction periods than their lenis counterparts and in all instances under review this parameter distinguished between the two series correctly. As far as voicing is concerned, the general tendency is that the lenis sounds have more voicing, but in some instances such as /θ/ (2) vs /ð/ (2) the differences are slight and depending on a speaker may not result in correct differentiation between the series.

To sum up, the analysis of fricatives suggests that voicing is mostly helpful in distinguishing between the fortis and lenis series, but friction length proved to be distinctive for all the fricatives under review both in word-initial and word-final position.

4. Discussion

In order to reconcile the results of the two studies summarised here, one ought to focus on two groups of features, i.e. voicing and other aspects of articulatory timing such as friction, aspiration or hold phase. The analysis of voicing suggests that it is not a feature that would always be sufficient to distinguish between the fortis and lenis series. In the case of plosives, it proves to be only incidentally decisive while it can also be inconclusive or misleading. In fricatives, voicing differences are clearly visible for /f/ and /v/, but are not fully conclusive for /θ/ and /ð/. Because of all that it seems that the phonetic voicing in articulation of the Welsh fricatives and plosives placed word-initially and word-finally in monosyllables is not distinctive for the fortis-lenis distinction.

The remaining three phonetic features, i.e. friction, aspiration and hold phase duration are analysed together, because they are all directly linked to the general duration of a sound. Voicing does not influence the general length of a consonant in such a direct way, because it occurs within aspiration or friction period. Moreover, there is a similarity between friction and aspiration, as aspiration is friction that appears after the release of the plosive. The analysis undertaken in the two studies summarised here suggests that these features are more reliable while distinguishing between the fortis and lenis sounds. Aspiration proves to be conclusive in all instances of plosive contrast under review and so does the hold phase. Friction length forms a stable paradigm throughout all instances of fricatives contrast under review. These observations support the point that the Welsh fortis-lenis laryngeal system is phonetically realised through articulatory timing and consonantal length. Depending on the manner of articulation the difference is visible either in aspiration and hold phase length in plosives or friction length in fricatives. Generally, there seems to be a pattern of the fortis obstruents being longer than their lenis counterparts.

5. Conclusions

This paper builds on the following findings of research on laryngeal systems: 1) for some languages, the fortis-lenis distinction is more appropriate than the voiceless-voiced distinction 2) the aforementioned phonological distinctions may be indicated by various phonetic features,

with the most common associations being the ones between phonetic voicing and [voice] and aspiration and [spread glottis] and 3) the same phonological feature may be responsible for the distinction between the series of sounds regardless of the manner of articulation, but its phonetic realisation may vary language-specifically.

Phonological research on the Welsh consonants suggests that it is one of the languages that employ the fortis-lenis contrast and that the phonetic voicing is of limited importance to it. The aim of this paper was, therefore, to verify it experimentally by studying Welsh obstruents. The plosives /p, b, t, d, k, g/ and fricatives /f, θ, v, ð, s, h/ were studied in monosyllabic lexemes in word-final and word-initial position. Aspiration, hold phase and voicing for plosives and friction and voicing for fricatives were measured in order to decide whether these factors are phonetically relevant in the fortis-lenis distinction of the Welsh obstruents.

The analysis confirmed that voicing is not fully conclusive when it comes to Welsh obstruents. The correlation was somewhat weak in plosives and slightly stronger in fricatives. Other aspects of articulatory timing proved decisive and formed clear correlations in all contrasts under review regardless of manner of articulation, position, prosodic position or place of articulation. These findings appear to support the fortis-lenis distinction of the Welsh obstruents and to disprove the voiceless-voiced one. The most important feature in the fortis-lenis distinction appears to be the articulatory timing – the fortis sounds seem to be longer than their lenis counterparts.

References

Asmus, S. and S. Grawunder. 2017. *Vowel length in Welsh monosyllables – Consonant-mutation dependence, Irish cognation and related theoretical problems. An acoustic study and its didactic implications.* Lampeter: The Edwin Mellen Press.

Asmus, S., and S. Grawunder. 2018. *Language-structuring consonant mutation(s) in Welsh (and Irish).* https://www.academia.edu/33509077/Language_structuring_consonant_mutation_s_in_Welsh_and_Irish_

Asmus, S., S. Jaworski, and M. Baran. 2019. Fortis-lenis of voiced-voiceless – Features of Welsh consonants. In S. Calhoun, P. Escudero, M. Tabain and P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019,* 3773–3776. Canberra: Australasian Speech Science and Technology Association Inc.

Ball, M. J. 1984. Phonetics for phonology. In M. J. Ball and G. E. Jones (eds.), *Welsh phonology. Selected readings,* 5–39. Cardiff: University of Wales Press.

Ball, M. J., and B. Williams. 2011. *Welsh phonetics.* Lewiston: The Edwin Mellen Press.

Baran, M. 2020. Friction and voicing length as phonetic correlates in the fortis-lenis distinction of the Welsh fricatives. *Proceedings of the Association of Celtic Students of Ireland and Britain Vol. VII*: 69–91.

Baran, M., and S. Asmus. 2019. A new perspective on Welsh and Irish consonant mutations. In S. Asmus and K. Jaworska-Biskup (eds.) *New perspectives on Modern Wales – Studies in Welsh language, literature and social politics,* 157–177. Cambridge: Cambridge Scholars Publishing.

Boersma, P., and D. Weeink. 2019. *Praat: Doing phonetics by computer [computer programme].* Version 6.1.08.

Butcher, A. 2004. Fortis-lenis revisited one more time: The aerodynamics of some oral stop contrasts in three continents. *Clinical linguistics and phonetics* 18: 547–557.

Chen, M. 2017. Vowel length variation as a function of the voicing of consonant environment. *Phonetica* 22: 129–159.

Cyran, E. 2011. Laryngeal realism and laryngeal relativism: Two voicing systems in Polish? *Studies in Polish Linguistics* 6: 45–80.

Hannahs, S. J. 2013. *The phonology of Welsh.* Oxford: Oxford University Press.

Hardcastle, W. J. 1973. Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics* 1: 263–272.

Hogan, J. T., and A. J. Rozsypal. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustic Society of America* 67: 1764–1771.

Honeybone, P. 2005. Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In M. van Oostendorp, and J. van de J. Weijer (eds.), *The internal organization of phonological segments,* 319–354. Berlin: Mouton de Gruyter.

Jaeger, J. J. 1983. The fortis/lenis question: Evidence from Zapotec and Jawon. *Journal of Phonetics* 11: 177–189.

Jakobson, R., G. M. Fant, and M. Halle. 1951. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge: MIT Press.

Jones, G. E. 1984. The distinctive vowels and consonants of Welsh. In M. J. Ball and G. E. Jones (eds.), *Welsh phonology. Selected readings,* 40–64. Cardiff: University of Wales Press.

Ladefoged, P. 1989. Representing phonetic structure. *UCLA Working Papers in Phonetics* 73: 1–79.

Lisker, L. 1986. 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29: 3–11.

Malécot, A. 1970. The lenis/fortis opposition: Its physiological parameters. *Journal of the Acoustic Society of America* 47: 1588–1592.

Odgen, R. 2009. *An introduction to English phonetics.* Edinburgh: Edinburgh University Press.

van Sluis, P. S. 2019. *Lenited voiceless stops in Middle Welsh: Phonology and orthography.* An unpublished PhD thesis. University of Wales.

Stetson, R. H. 1951. *Motor phonetics.* Amsterdam: North Holland.

Stevens, K. N. 1998. *Acoustic phonetics.* Cambridge: MIT Press.

Trask, R. L. 1996. *A dictionary of phonetics and phonology.* New York: Routledge.

Word-based largest chunks for Agreement Groups processing: Cross-linguistic observations

László Drienkó

SZSZC-Jáky Székesfehérvár, Hungary

Abstract

The present study reports results from a series of computer experiments seeking to combine word-based Largest Chunk (LCh) segmentation and Agreement Groups (AG) sequence processing. The AG model is based on groups of similar utterances that enable combinatorial mapping of novel utterances. LCh segmentation is concerned with cognitive text segmentation, i.e. with detecting word boundaries in a sequence of linguistic symbols. Our observations are based on the text of *Le petit prince* (*The little prince*) by Antoine de Saint-Exupéry in three languages: French, English, and Hungarian. The data suggest that word-based LCh segmentation is not very efficient with respect to utterance boundaries, however, it can provide useful word combinations for AG processing. Typological differences between the languages are also reflected in the results.

Keywords: Cognitive computer modelling; segmentation; syntactic processing; language acquisition

1. Introduction

The AG language processing model as proposed in Drienkó (2014) is a usage-based distributional framework where groups of utterances are formed according to the distribution of words in a given corpus. Distributional linguistic research can be traced back at least to Harris (1951, 1952). In Harris's work the contexts, or environments, of a linguistic element were used to determine the distribution of the element in question. The contexts of words or phrases are particularly helpful in categorisation research based on cluster analysis (e.g. Kiss 1973, Redington *et al.* 1998, Finch *et al.* 1995), where context is typically formalised by context vectors. In Mintz (2003) a context, or *frame,* is provided by words that immediately precede or follow a given target element and a *frequent frame* is a context occurring with a frequency above an arbitrary threshold value. Weisleder and Waxman (2010) propose the utilization of *end-frames* with utterance-end information for categorisation. St. Clair *et al.* (2010) suggest that *flexible frames*, with bigram information from contexts, are more suited for categorising than only frequent frames. Item-based phrases in language acquisition research, as framed by words in initial positions, constitute a specific type of context (e.g. Cameron-Faulkner *et al.* 2003, Stoll

et al. 2009). AGs may be viewed as superimpositions of such contextual frames. According to Wang and Mintz (2010) "grammatical relations between words are more consistent in individual frequent frames than in bigrams" and "words within a frequent frame are especially "close" syntactically" (6, 8). Such views accord with our assumption that the "agreement relations" encoded in AGs represent syntactical/grammatical relations.

Early work on speech segmentation is exemplified by Harris (1955). His research focussed on statistical characteristics of language, fundamentally on successor frequencies, which he used for predicting word or morpheme boundaries. As documented by Saffran, Aslin and Newport (1996), infants may indeed be supported by statistical characteristics of speech in acquiring language. Research on speech segmentation has also demonstrated that several lexical and sub-lexical language-related cues play an important part in language acquisition (e.g. Mattys, White and Melhorn 2005). These cues can be utilised by various segmentation strategies. Metrical segmentation (Cutler and Carter 1987; Cutler and Norris 1988), for instance, is based on the distribution of strong and weak syllables. Also, infants can rely on stress patterns (Thiessen and Saffran 2007), or lengthening of speech sounds and/or rising in acoustic frequency (Bagou, Fougeron, and Frauenfelder 2002) for segmenting language. The LCh segmentation strategy does not employ such cues. It only needs information on the succession of linguistic elements in a particular text.

The structure of this paper is as follows: Sections 1.1 and 1.2 provide a short introduction to AGs and LCh segmentation. Section 1.3 sets the scene for the experiments by considering the issue of the possible combination of word-based LCh segmentation with AG processing. In Section 2, we present our empirical findings. In Section 3, we discuss the significance of the results with respect to linguistic modelling. Section 4 contains some concluding remarks.

1.1. Agreement Groups

The AG model of language processing is a usage-based distributional framework operating with memorised groups of similar utterances, and cognitive mapping mechanisms. Thus a collection of familiar/known utterances enables the processing of novel word sequences. Formally, an AG can be regarded as a hypothetical table for concatenating linguistic units, where columns in the table represent (agreement) categories, and any element (word) in a column can be concatenated with any other in the next column.

The idea of *agreement groups* and *agreement groups coverage* was presented in a series of works as a distributional approach to modelling linguistic processing. Drienkó (2014) showed that agreement groups, i.e. groups of 2–5 word long utterances differing from a base utterance in only one word, can account for a certain percent of novel utterances of English mother-child speech, may facilitate categorisation (lexical/syntactic, semantic), and might serve as a basis for 'real' agreement relations. The findings were confirmed cross-linguistically by Hungarian and Spanish data in Drienkó (2013a). For the processing of longer utterances, the notion of *coverage* was introduced in Drienkó (2013b, 2015, 2016b). The coverage apparatus seeks to identify 2–5 word long fragments of an input utterance and map them onto AGs. By applying the AG coverage method to mother-child speech (Anne sessions, Manchester corpus: Theakston *et al.*

2001) from the CHILDES corpora, (MacWhinney 2000), it was found that the continuous and the discontinuous cases yielded, respectively, 78% and 83% average coverage values.

The essence of the AG approach lies in forming groups differing in only one word from a given utterance. In fact, each utterance of the training set has its own group. For instance, the training corpus (1) yields the AGs under (2).[1]

(1)   the dog   the cat   big dog   big cat   white dog   the big dog
       the white dog   the big cat   the dog laughs   the cat laughs   the dog cries   cat laughs   dog laughs

(2)   G1:   G2:   G3:   G4:   G5:   G6:
       the dog   the cat   big dog   big cat   white dog   the big dog
       the cat   the dog   big cat   big dog   big dog   the big cat
       big dog   big cat   white dog   the cat   the dog   the white dog
       white dog   the dog

       G7:   G8:   G9:   G10:   G11:   G12:   G13:
       the white dog   the big cat   the dog laughs   the cat laughs   the dog cries   cat laughs   dog laughs
       the big dog   the big dog   the dog cries   the dog laughs   the dog laughs   dog laughs   cat laughs
                                        the cat laughs

We think of groups as hypothetical tables as defined by the utterance length for the group (number of columns in the table), and the maximum number of words occurring in an utterance position (number of rows). We say that an utterance is compatible with a group (i.e. can be mapped on a group) if it can be obtained by choosing words from the subsequent columns of the corresponding hypothetical table. Although the novel utterance *white cat*, e.g., is not an utterance of the training set, it can be mapped on the *the dog* group, G1, or on the *big dog* group, G3. The assignment of 'agreement categories' is done with reference to groups and utterance positions, cf. (3). Categories *G9_3* and *G11_3* for *cries*, for instance, indicate that the word occurs in Group 9 and Group 11 in the third word position within the corresponding utterances.

(3)   the: G1_1, G2_1, G3_1, G4_1, G5_1, G6_1, G7_1, G8_1, G9_1, G10_1, G11_1
       big: G1_1, G2_1, G3_1, G4_1, G5_1, G6_2, G7_2, G8_2,
       white: G1_1, G3_1, G5_1, G6_2, G7_2
       dog: G1_2, G2_2, G3_2, G4_2, G5_2, G6_3, G7_3, G8_3, G9_2, G10_2, G11_2, G12_1, G13_1
       cat: G1_2, G2_2, G3_2, G4_2, G6_3, G8_3, G9_2, G10_2, G12_1, G13_1
       laughs: G9_3, G10_3, G11_3, G12_2, G13_2
       cries: G9_3, G11_3,

The COVERAGE STRUCTURE of an utterance is a tabular visualisation of a configuration of AGs onto which the fragments of the utterance in question can be mapped. For instance, Table 1 shows the possible fragments that can cover sentence *the big white dog laughs*. In Table 2 the words are represented by their agreement categories directly indicating which groups are involved.

---

[1] Examples from Drienkó (2017a).

Table 1: *Schematic coverage structure for* the big white dog laughs

the	big	white	dog	laughs
			dog	laughs
the			dog	
	big		dog	
		white	dog	
the	big		dog	
the		white	dog	
the			dog	laughs

Table 2: *Coverage structure with category information for* the big white dog laughs

the	big	white	dog	laughs
			G12_1	G12_2
			G13_1	G13_2
G1_1			G1_2	
...			...	
G5_1			G5_2	
	G1_1		G1_2	
	...		...	
	G5_1		G5_2	
		G1_1	G1_2	
		G3_1	G3_2	
		G5_1	G5_2	
G6_1	G6_2		G6_3	
G7_1	G7_2		G7_3	
G8_1	G8_2		G8_3	
G6_1		G6_2	G6_3	
G7_1		G7_2	G7_3	
G9_1			G9_2	G9_3
G10_1			G10_2	G10_3
G11_1			G11_2	G11_3

The AG model assumes two basic levels of linguistic processing. The first level corresponds to direct mappings onto AGs for processing holophrases, shorter utterances, or "formulaic" expressions. The second level requires more computational effort since firstly legal (i.e. AG-compatible) fragments have to be found (Level 1 operation), then an optimal combination of fragments must be selected in order to effect grammaticality. This duality is reflected in the coverage structures of utterances. Further dualistic properties of the AG framework are communicated in Drienkó (2018a, 2020) along with contact points for research on cognitive-linguistic processing including generalisation, categorisation, a semantic/syntactic categorical interpretation of the *less-is-more* principle of Newport (1990) and its relationship to U-shaped learning (Strauss, 1982) and vocabulary spurt (e.g. Ganger and Brent 2004), parallelisms with the dual-process model of Van Lancker Sidtis (2009), lateralization of formulaic and analytical speech (e.g. Sidtis, Sidtis, Dhawan, and Eidelberg 2018), neurolinguistic processing (Bahlmann *et al.* 2006), and the processing of complex linguistic structures such as long-distance dependencies, crossing dependencies, or embeddings (cf. also Drienkó 2016b).

1.2. Largest-Chunk segmentation

The LCh segmentation algorithm as proposed in Drienkó (2016a) searches for a succession of language chunks in an unsegmented sequence of linguistic symbols, which chunks are locally maximal in length and occur minimally twice in the whole sequence. To quantify the empirical results, four precision values are computed: INFERENCE PRECISION (IP), ALIGNMENT PRECISION (AP), REDUNDANCY (R), and BOUNDARY VARIABILITY (BV). As an immediate example, consider the toy corpus {*mary is, mary it*} consisting of two utterances. When the basic segmentation units are the characters of the text, the LCh algorithm outputs the segments *maryi*, *s*, *maryi*, and *t* as in (4). Since 2 boundaries are correct of all the 4 inferred boundaries – viz. the boundaries after *s* and *t* –, IP is 2/4=0.5. Note that *IP=cib/aib*, i.e. the number of correctly inferred/inserted boundaries (*cib*) divided by the number of all inferred/inserted boundaries (*aib*).

(4)   maryismaryit → maryi s maryi t

When segmentation is based on syllables, we expect higher precision since boundaries cannot be erroneously inferred syllable-internally. The LCh segments for our example corpus {*mary is, mary it*} would be *ma-ry-*, *is-*, *ma-ry-*, and *it-*, cf. (5). Now IP=4/4=100%, since each of the four original boundaries is inferred correctly.

(5)   ma-ry-is-ma-ry-it- → ma-ry- is- ma-ry- it-

In the cross-linguistic analysis of Drienkó (2017b), letter/character-based LCh segmentation was applied to utterances from English, Hungarian, Mandarin, and Spanish. The analysis yielded a 53% – 66% IP range, averaging 59%. Drienkó (2018b) examined how the precision values are affected when syllables are the basic segmentation units. It was found that syllable-based LCh segmentation results in considerably higher IP values, within an interval of 77%–95%, averaging 86%.

The LCh segmentation strategy may be compatible with the approach of Peters (1983) where a key role in language acquisition is played by segmenting and fusing linguistic chunks extracted from a continuous stream of speech. The LCh segmentation results might also suggest an analogy with the less-is-more interpretation of the data in Newport (1990), i.e. with the claim that certain cognitive skills may develop at the expense of others. In our case, boundary inference is more efficient when the processing of syllable structure (characters) is suppressed, i.e. when the syllable is taken to be the basic segmentation unit. Although the LCh strategy does not require cues like, for instance, metrical features, or stress patterns, it may be compatible with other cognitive strategies, and it can be aided by cognitive cues. In Drienkó (2018c) it was reported that LCh segmentation is enhanced by utterance boundary information, which fact is congruent with findings from infant word segmentation research. Indeed, the Edge Hypothesis of Seidl and Johnson (2006), in particular, assumes that utterance boundaries may provide an important cue in segmentation.

1.3. *Word-based Largest Chunks for Agreement Groups*

The AG model tacitly assumes that utterance boundaries are readily available to the language learner, i.e. the training corpus consists of utterances with their well-defined boundaries. However, this is an over-optimistic attitude with regards to real-life natural language acquisition. The learner of a language is normally exposed to continuous speech without evident boundary markers. Previous research findings (Drienkó 2017b, 2018b) indicated that word boundaries can be detected via the Largest Chunk strategy with fairly high precision, especially for the syllable-based case. Assuming, then, that the language learner has a tool for detecting word boundaries (e.g. syllable-based LCh segmentation) it might be insightful to examine, as a next step, how the LCh segmentation strategy can be useful when the word is taken to be the basic textual unit. It might be expected that the strategy can detect reoccurring word combinations corresponding to phrases and utterances. These "phrases" (or rather speech fragments), in turn, could be input to the group formation algorithm of the AG model. Finally, the resultant body of AGs could condition a mapping mechanism for novel word sequences. Thus there could be a cognitive computer model for the emergence of language, basically building on two cognitive capacities, LCh segmentation, and AG formation together with the concomitant mapping mechanisms.

The present study reports results from a series of experiments seeking to combine word-based LCh segmentation with the AG utterance processing apparatus. In the experiments, first, the input corpus of utterances was transformed into a sequence of words by deleting punctuation symbols, i.e. utterance boundaries, and the resultant word sequence was segmented by the LCh segmentation algorithm.[2] In the next phase, the collection of word combinations (largest chunks) obtained in the first stage was used for producing AGs. Finally, the resultant AGs were used for mapping utterances of a novel section (test set) of the original corpus, i.e. for testing coverage. For computational reasons, we decided to include utterance boundaries in the test set. This means that our results quantitatively underestimate the coverage potential of the model in that word combinations possibly spanning utterance boundaries are ignored.

2. The experiments

Our observations are based on the text of *Le petit prince* (*The little prince*) by Antoine de Saint-Exupéry (1943a,b,c) in three languages: French, English, and Hungarian. The book contains 27 chapters. For each language, we utilised Chapters 1–26 for producing text segments whereas Chapter 27 was used for testing the coverage potential of AGs. In the segmentation phase, the text was divided into five subparts – Chapters 1–5, 6–10, 11–15, 16–21, and 22–26 – and each subpart was segmented separately. However, for a given language, segments from all the five subtexts were considered. For instance, in Experiment 1 the first collection of segments came

---

[2] Since the texts contain long and complex sentences, we chose to identify boundaries demarked by punctuation symbols including e.g. the comma, colon, or brackets, with utterance boundaries. That means that in the present study the term 'utterance boundary' should rather be understood as also subsuming clause or phrase boundaries besides sentence boundaries.

from Chapters 1–5 of the French text, the second collection from Chapters 6–10 etc., and the segments of all the five collections were used to form AGs. Coverage was then tested on Chapter 27. The same holds for Experiments 2 and 3 with the English and Hungarian version of the book, respectively.

2.1. Experiment 1: French

In Experiment 1 the LCh segments were obtained from Chapters 1 through 26 of the original French text. Overall, there were 9665 segment tokens, 3522 types, provided by the 5 datasets.

Table 3 shows the precision metrics for the segmentation procedure. Note that *IP=cib/aib*, i.e. the number of correctly inferred/inserted boundaries, *cib*, divided by the number of all inferred/inserted boundaries, *aib*; *R=aib/acb*, i.e. the number of all inferred/inserted boundaries, *aib*, divided by the number of all correct, original, boundaries, *acb*; *AP=cib/acb*, i.e. the number of correctly inferred/inserted boundaries divided by the number of all the correct boundaries; and *BV* stands for the average distance between an inferred boundary and the nearest correct one, measured in characters.

Table 3: *Segmentation precision and coverage results for Experiment 1*

Le petit prince						
	1–5	6–10	11–15	16–21	22–26	Average
IP	0.13	0.15	0.18	0.16	0.16	0.16
R	7.11	6.16	4.92	5.41	5.55	5.83
AP	0.94	0.91	0.89	0.89	0.88	0.90
BV	20.91	18.57	15.76	17.07	17.13	17.89
Average coverage (cont.)	0.58					
Average coverage (discont.)	0.66					

Of all the 3522 segment types 1112 were multiword segments containing at most five words. These 1112 two-to-five-word-long segments were used for the formation of AGs. They contained 585 word types. Since each segment had its own group, there were 1112 AGs. The text of Chapter 27 was used for testing the coverage potential of this 1112-group AG system. The chapter consists of 37 sentences. In order to minimise computational costs sentence boundaries were retained, as well as boundaries demarcated by other punctuation symbols, e.g. commas and colons. One-word utterances were excluded from the analysis as meaningless for syntactic processing since AG-utterances minimally consist of two words. The test set eventually contained 70 text fragments which were input to the coverage evaluation procedure. By coverage we mean the percentage of utterance positions covered by at least one fragment mappable on some AG. For instance, assuming that utterance fragments *the dog, clever creature,* and *is a creature* can be mapped on some AGs, the coverage value for utterance '*the dog is a clever creature*' is 4/6 = 67% since four of the six utterance positions are covered by fragments *the dog,* and *clever creature.* This is the non-discontinuous case. In the discontinuous case, we would say that coverage is 100%, as the *is* and *a* positions of the sentence could be covered

discontinuously by *is a creature*, cf. Tables 4 and 5 displaying the continuous and discontinuous coverage structure for the utterance *the dog is a clever creature*.

Table 4: *Coverage structure for 'the dog is a clever creature' (continuous fragments only)*

the	dog	is	a	clever	creature
the	dog				
				clever	creature

Table 5: *Coverage structure for 'the dog is a clever creature' (discontinuous fragments allowed)*

the	dog	is	a	clever	creature
the	dog				
				clever	creature
		is	a		creature

Via dividing the sum of the coverage values for each utterance in the test set by the number of utterances in the test we obtain average coverage. The average coverage value from Experiment 1 was 40.36 / 70 = 57.6% for the continuous case and 46.44 / 70 = 66.3% for the discontinuous case, cf. Table 3.

2.2. Experiment 2: English

In Experiment 2 the LCh segments came from Chapters 1 through 26 of the English translation of the book. Overall, there were 9316 segment tokens, 3046 types, provided by the 5 datasets. Table 6 shows the precision metrics for the segmentation procedure.

Table 6: *Segmentation precision and coverage results for Experiment 2*

The Little Prince						
	1–5	6–10	11–15	16–21	22–26	Average
IP	0.12	0.15	0.18	0.16	0.16	0.15
R	6.99	6.07	4.83	5.44	5.32	5.73
AP	0.87	0.89	0.87	0.89	0.85	0.87
BV	19.18	16.77	14.47	16.56	15.77	16.55
Average coverage (cont.)	0.58					
Average covergae (discont.)	0.67					

Of all the 3046 segment types 1140 were multiword segments containing at most five words. These 1140 two-to-five-word-long segments were used for the formation of AGs. They contained 483 word types. The text of Chapter 27 was used for testing the coverage potential of the 1140-group AG system. Due to the retention of punctuation-effected boundaries, the 37 sentences of the chapter were represented by 66 text fragments. The average coverage value in Experiment 2 was 38.45 / 66 = 58.3% for the continuous case and 44.23 / 66 = 67% for the discontinuous case, cf. Table 6.

2.3. Experiment 3: Hungarian

In Experiment 3 the LCh segments were provided by Chapters 1 through 26 of the Hungarian translation of the book. Overall, we obtained 9260 segment tokens, 4053 types from the 5 datasets. Table 7 shows the precision metrics for the segmentation procedure.

Table 7: Segmentation precision and coverage results for Experiment 3

Kis herceg						
	1–5	6–10	11–15	16–21	22–26	Average
IP	0.12	0.18	0.22	0.20	0.19	0.18
R	7.59	5.36	4.18	4.29	4.79	5.24
AP	0.94	0.97	0.92	0.88	0.95	0.93
BV	23.86	17.32	12.81	14.62	14.51	16.62
Average coverage (cont.)	0.20					
Average coverage (discont.)	0.28					

Of all the 4053 segment types 533 were multiword segments containing at most five words. The 533 two-to-five-word-long segments were used for the formation of AGs. They contained 416 word types. Chapter 27 was used for testing the coverage potential of the 533-group AG system. Due to the retention of punctuation-effected boundaries, the 37 sentences of the chapter were represented by 84 text fragments. The average coverage value in Experiment 3 was 16.51 / 84 = 19.6% for the continuous case and 23.57/ 84 = 28.06% for the discontinuous case, cf. Table 7. Table 8 presents the average results from all the three experiments.

Table 8: Overall average segmentation precision and coverage results

	PP	LP	KH	Average
Average IP	0.16	0.15	0.18	0.16
Average R	5.83	5.73	5.24	5.6
Average AP	0.90	0.87	0.93	0.9
Average BV	17.89	16.55	16.62	17.02
Average covergae (cont.)	0.58	0.58	0.20	0.45
Average covergae (discont.)	0.66	0.67	0.28	0.54

3. Discussion

The Inference Precision (IP) values show that the number of correctly inferred boundaries as compared to the number of all inferred boundaries is rather low, 16%, on average. This suggests that the LCh segmentation mechanism, as compared to previous results (Drienkó 2017b, 2018b), is not very efficient when words are the basic segmentation units and utterances are the target sequences, i.e. utterance boundaries are to be inferred. However, the other precision values reveal further features of the LCh strategy that make it capable of providing useable word combinations for syntactic processing. As the 90% average Alignment Precision (AP) value indicates, almost all of the utterance boundaries are correctly identified. The high AP value is

achieved via inserting extra boundaries. The 5.6 average Redundancy value shows that more than five times as many boundaries are inferred as would be strictly necessary to identify the original utterances. The extraneous boundaries are incorrect with respect to utterance edges. Nevertheless, they delineate reoccurring word sequences that can be used as building blocks for utterances. As reflected in the coverage values, such building blocks, or "phrases" can account for, on average, ca. 50% of the text.

For each language, the coverage value is higher when discontinuous fragments are permitted in processing. This fact echoes the findings in Drienkó (2015) claiming that discontinuous fragments in the coverage mechanism enhance the coverage potential of the AG model.

The data also reflect typological differences between the languages involved in the experiments. While the segmentation metrics are remarkably similar across languages, the 20% and 28% coverage values for Hungarian stand in clear contrast to the corresponding values for French and English, well over 50%, cf. Table 8. Since Hungarian is a highly inflectional language with relatively free word order, words and utterances are less likely to reoccur in the same form as in English or French. As repetitions are vital for LCh segmentation, just as similarity of word combinations is a key determinant in the formation of AGs, it can be expected that languages with a high degree of word-form variation and/or variable word order require more extensive training input in order to achieve the same level of efficiency of AGs. In other words, while Chapters 1–26 of the English and French texts provide enough similar segments for the resultant AGs to achieve relatively high coverage, that is not the case for Hungarian. The French and English training texts provided 1112 and 1140 word combinations, i.e. AGs, respectively. For Hungarian, the number of AGs was 533, ca. half the number of AGs for either French or English. Note that the differences seem to suggest a correlation between AG space and coverage. The same line of argumentation may be valid in explaining why the coverage values are slightly higher for English (58.3%, 67%) than for French (57.6%, 66.3%), assuming that French has richer inflectional morphology and freer word order. The English-French-Hungarian ranking also mirrors the morphological complexity of the languages as assumed to negatively correlate with the number of native speakers (e.g. Koplenig 2019): English, the least inflectional language, is spoken by the most native speakers, whereas the most inflectional, Hungarian, has rather few native speakers.

Boundary Variability (BV) is rather high, the average value is 17.02. This means that, on average, the distance between an inferred boundary and the nearest true one is about 17 characters. Via dividing BV by the average word length for the particular language we get an estimation of BV_{wo}, i.e. 'Boundary Variability measured in words'. Table 9 shows the average BV_{wo} values calculated as BV/WL, where WL stands for '(average) word length'. The data reveal that, across the languages, the 17.02 character-based average BV corresponds to an average distance of 3.6 words from the nearest correct boundary. Again, dissociation can be observed in terms of language types. Hungarian can express grammatical dependencies within a sentence inflectionally. For instance, a single verb can refer to the subject and/or the object. In English and French, explicit parts of speech are needed for the subject or the object. Such facts suggest that Hungarian needs fewer albeit longer words to build a sentence or utterance. Fewer words in utterances, in turn, imply less chance to err in boundary inference (cf. also the IP and AP

values in Table 8). For instance, with a two-word-long utterance there is only one possibility to make an error, i.e. when a boundary is inserted between the two words. The distance of the incorrect boundary from either the boundary before the first word or the boundary after the second word is then one word. With a four-word-long utterance, an incorrectly inserted boundary in the middle would be two words away from either the left or the right correct boundary. Thus growing utterance length involves utterance positions that can possibly increase BV_{wo}. Consequently, the lower BV_{wo} value for Hungarian than for either English or French, and the lower BV_{wo} value for French than for English might ultimately be ascribed to morphological differences affecting utterance length.

Table 9: *Average BV, word length (WL) and $BV_{wo} = BV/WL$*

	PP	LP	KH	Average
Average BV	17.89	16.55	16.62	17.02
Average WL	4.6	4.2	5.3	4.7
Average BV_{wo}	3.88	3.94	3.13	3.6

4. Conclusions

The primary purpose of the present study was to investigate, cross-linguistically, the viability of combining word-based LCh segmentation with AG processing. We reported empirical results from experiments with the text of *The Little Prince*. It was found that word-based segmentation is not particularly efficient for inferring utterance boundaries, IP is ca. 16%. However, the majority of utterance boundaries can be reconstructed, AP≈90%, by way of inserting redundant boundaries, R≈5.6. The resultant abundance of segments, in turn, conditions the emergence of utterance components, or building blocks, that can be organised into AGs. Thus LCh segmentation provides useable word combinations for syntactic processing. As reflected in the coverage values, such building blocks, or "phrases" can account for about 50% of the test texts, on average, rendering our approach a promising processing framework. The data also highlight typological differences between the languages involved.

Our findings may be considered preliminary and need further validation against more extensive corpora. One step in that direction could be the analysis in Drienkó (in review) based on English mother-child utterances, with coverage over 80%. If it turns out to be adequately supported by empirical data, the 'LCh+AG' approach can offer a footing for establishing a usage-based model/theory of the emergence of language capacities built around two fundamental cognitive strategies, LCh segmentation and AG formation. The model might also be compatible with traditions in language acquisition research. Erickson and Thiessen (2015), e.g., conceptualise statistical learning as consisting of two major processes, Extraction and Integration. Extraction refers to statistical chunking whereas Integration involves similarity-weighted aggregation over chunks. Our LCh segments implicitly reflect the statistical-distributional structure of a sequence of symbols (words, in the present work) whereas the grouping of the segments into AGs is dictated by their distributional similarity-statistics.

References

Bagou, O., C. Fougeron, and U. H. Frauenfelder. 2002. Contribution of prosody to the segmentation and storage of "Words" in the acquisition of a new mini-language. *Speech Prosody 2002*, Aix-en-Provence, France, April 11–13, 2002.

Bahlmann, G., and A. D. Friederici. (2006). Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience* 18(11): 1829–1842.

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

Cutler, A., and D. M. Carter. 1987. The predominance of strong initial syllables in English vocabulary. *Computer Speech and Language* 2: 133–142.

Cutler, A. and D. G. Norris. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113–121.

de Saint-Exupéry, A. 1943a. *Le petit prince*. Édition du groupe "Ebooks libres et gratuits". Retrieved from https://www.ebooksgratuits.com/pdf/st_exupery_le_petit_prince.pdf.

de Saint-Exupéry, A. 1943b. *The little prince*. English translation by Jeff Mcneill. TheVirtualLibrary.org. Retrieved from https://thevirtuallibrary.org/index.php/en/?option=com_djclassifieds&format=raw&view=download&task=download&fid=14329.

de Saint-Exupéry, A. 1943c. *A kis herceg*. Hungarian translation by György Rónay. Retrieved from http://mek.oszk.hu/00300/00384/00384.pdf.

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

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

Drienkó, L. 2014. Agreement groups analysis of mother-child discourse. In G. Rundblad, A. Tytus, O. Knapton, and C. Tang (eds.), *Selected Papers from the 4th UK Cognitive Linguistics Conference*, 52–67. London: UK Cognitive Linguistics Association. Retrieved from http://www.uk-cla.org.uk/proceedings/volume_2_36/36-32.

Drienkó, L. 2015. Discontinuous coverage of English mother-child speech. Talk presented at the *Budapest Linguistics Conference,* Budapest, Hungary, 18–20 June, 2015.

Drienkó, L. 2016a. Discovering utterance fragment boundaries in small unsegmented texts. In A. Takács, V. Varga, and V. Vincze (eds.), *XII. Magyar Számítógépes Nyelvészeti Konferencia* (*12th Hungarian Computational Linguistics Conference*), 273–281. Retrieved from https://rgai.inf.u-szeged.hu/sites/rgai.sed.hu/files/MSZNY2016_web_ISO_B5.pdf.

Drienkó, L. 2016b. Agreement groups coverage of English mother-child utterances for modelling linguistic generalisations. *Journal of Child Language Acquisition and Development – JCLAD* 4(3): 113–158. Retrieved from http://jclad.science-res.com/archives_full_issu/Vol%204%20issue%203%20FULL%20ISSUE.pdf.

Drienkó, L. 2017a. Agreement groups processing of context-free utterances: Coverage, structural precision, and category information Talk presented at the *2nd Budapest Linguistics Conference,* 1–3 June 2017, Budapest, Hungary.

Drienkó, L. 2017b. Largest chunks as short text segmentation strategy: A cross-linguistic study. In A. Wallington, A. Foltz, and J. Ryan (eds.), *Selected Papers from the 6th UK Cognitive Linguistics Conference*, 273–292. The UK Cognitive Linguistics Association. Retrieved from http://www.uk-cla.org.uk/files/downloads/15_drienko_273_292.pdf.

Drienkó, L. 2018a. Agreement groups and dualistic syntactic processing. Talk presented at the "*One Brain – Two Grammars? Examining dualistic approaches to language and cognition*" international workshop, 1–2 March 2018, Rostock, Germany. Retrieved from https://independent.academia.edu/LaszloDrienko/Conference-Presentations.

Drienkó, L. 2018b. Largest-Chunk strategy for syllable-based segmentation. *Language and Cognition* 10(3), 391–407.

Drienkó, L. 2018c. The effects of utterance-boundary information on Largest-Chunk segmentation. Talk presented at the *20th Summer School of Psycholinguistics,* Balatonalmádi, Hungary, 10–14 June, 2018.

Drienkó, L. (in review). Largest-chunking and group formation: two basic strategies for a cognitive model of linguistic processing.

Drienkó, L. 2020. Agreement Groups and dualistic syntactic processing. In A. Haselow, and G. Kaltenböck (eds.), *Grammar and cognition: Dualistic models of language structure and language processing*, 310–354. John Benjamins Publishing Company.

Erickson, L. C., and E. D. Thiessen. 2015. Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review* 37: 66–108.

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

Ganger, J., and M. R.Brent. 2004. Reexamining the Vocabulary Spurt. *Developmental Psychology* 40(4): 621–632.

Harris, Z. S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

Harris, Z. S. 1952. Discourse analysis. *Language* 28(1): 1–30.

Harris, Z. S. 1955. From phoneme to morpheme. *Language* 31: 190–222.

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

Koplenig, A. 2019. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science* 6: 181274. Retrieved from https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.181274.

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

Mattys, S. L, L. White, J. F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134(4): 477–500.

Mintz, T. H. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90(1): 91–117.

Newport, E. L. 1990. Maturational constraints on language learning. *Cognitive Science* 14: 11–28.

Peters, A. 1983. *The units of language acquisition*. Cambridge: Cambridge University Press.

Redington, M., N. Chater, and S. Finch. 1998. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22(4): 425–469.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294): 1926–1928.

Seidl, A. and E. K. Johnson. 2006. Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science* 9: 565–573.

Sidtis, J. J., D. Van Lancker Sidtis, V. Dhawan, and D. Eidelberg. 2018. Switching Language Modes: Complementary Brain Patterns for Formulaic and Propositional Language. *Brain connectivity* 8(3): 189–196.

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

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

Strauss, S. 1982. Ancestral and descendent behaviours: The case of U-shaped behavioural growth. In T. G. Bever (ed.), *Regressions in mental development: Basic phenomena and theories*, 191–220. Hillsdale, NJ: Lawrence Erlbaum Associate, Inc.

Theakston, A. L., E. V. Lieven, J. M. Pine, and C. F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*. 28(1):127–52.

Thiessen, E. D., and J. R. Saffran. 2007. Learning to Learn: Infants' Acquisition of Stress-Based Strategies for Word Segmentation. *Language Learning And Development* 3(1): 73–100.

Van Lancker Sidtis, D. 2009. Formulaic and novel language in a 'dual process' model of language competence: Evidence from surveys, speech samples, and schemata. In R. L. Corrigan, E. A. Moravcsik, H. Ouali, and K. M. Wheatley (eds.), *Formulaic Language: Volume 2. Acquisition, loss, psychological reality, functional applications*, 151–176. Amsterdam: Benjamins Publishing Co.

Wang, H., and T. H. Mintz. 2010. From Linear Sequences to Abstract Structures: Distributional Information in Infant-direct Speech. In J. Chandlee, K. Franich, K. Iserman, and L. Keil (eds.), *Boston University Conference on Language Development 34 Online Proceedings Supplement.* Somerville, MA: Cascadilla Press. Retrieved from http://www.bu.edu/bucld/proceedings/supplement/vol34/.

Weisleder, A, and S. R. Waxman, 2010. What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language* 37(5): 1089–108.

Similarities and differences between two Hungarian particles for also: *szintén* and *is*

Judit Farkas, Bettina Futó, Aliz Huszics, Judit Kleiber, Mónika Dóla
and Gábor Alberti
University of Pécs, Hungary

Abstract

The paper provides a comparative analysis of the syntax, semantics and pragmatics of two Hungarian particles with the same logical core meaning also: *is* and *szintén*. The analysis yields important theoretical implications since it demonstrates how two particles sharing the same logical-propositional/truth-functional core meaning can expand into two different markers. In discourse, *is* acts as an intensional/metacognitive *pragmatic marker* in the sense as proposed by Aijmer *et al.* (2006), while *szintén* functions as a coherence-signaling *discourse marker*. The two particles share certain syntactic-semantic properties: neither of them can be followed by a topic, they both have distributive meaning, and both of them can pertain to the noun phrase that they immediately follow, as well as to ordered n-tuples of noun phrases. However, there are also syntactic and pragmasemantic differences between them. Namely, their ordered n-tuples have different word orders; *is* can function as a pragmatic marker while *szintén* cannot; *szintén* can appear as a separate clause, while *is* cannot (this is presumably related to the fact that *szintén* can be stressed, while *is* is obligatorily unstressed); and finally, *szintén* can have a peculiar discourse-preserving function. We explain the syntactic differences between the two particles using the *partial spell-out* technique of minimalist generative syntacticians (first applied to Hungarian by Surányi 2009), and the Cinque-hierarchy-based approach to Hungarian sentence- and predicate-adverbials (Surányi 2008). We account for the pragmasemantic properties of the pragmatic-marker variant of *is* in the formal representational dynamic theory of interpretation called ℜeALIS, already presented in the LingBaW series (Alberti *et al.* 2016, Kleiber and Alberti 2017, Viszket *et al.* 2019).

Keywords: particles, distributive meaning, Hungarian, pragmatic and discourse markers

1. Introduction

The paper analyzes the similarities and differences between two Hungarian particles with the same lexical core meaning ALSO: *is* [ɪsh] and *szintén* [sɪnteɪn].

Is as a particle has had a long history in Hungarian generative literature, with such seminal papers clarifying its basic properties as Brody (1990) and É. Kiss (1992, 2002). There is a lack,

however, of (widely known and accepted) research on *szintén*.[1] Since both particles share the meaning ALSO, it seems only reasonable to start the investigation of *szintén* in the light of *is*. Furthermore, their parallel comparative investigation can provide a useful contribution to the universal research into polysemic systems around multifunctional phonetic forms (Fischer 2006: 13–14), with special regards to the subtle distinction between discourse markers and pragmatic markers as proposed by Aijmer *et al.* (2006). In this terminology, a pragmatic marker is a word or expression that does not contribute to the propositional, truth-functional content of an utterance, while a discourse marker signals coherence relations. In this respect, the major theoretical implication of the parallel analysis of the two Hungarian particles is that it demonstrates a case where two particles sharing the same logical-propositional/truth-functional core meaning (ALSO) expand into two different markers: *is* into an intensional/metacognitive *pragmatic marker*, and *szintén* into a coherence-signaling *discourse marker*.

Section 2 provides an overview of the (primarily truth-functional) similarities between the two particles in question. Section 3 presents the (dominantly syntactic) differences between them. Section 4 offers their syntactic representations, including those that capture the coherence-signaling discourse-marker character of *szintén*. Section 5 is devoted to the demonstration of the pragmatic-marker function of *is*, through presenting a few types of its pragmatic use. The paper concludes with a summary of the main findings (Section 6).

2. Similarities between *is* and *szintén*

This section provides an overview of the properties of *is* that also hold true for *szintén*, to varying degrees.

2.1. The distributive nature of is and szintén

Is is a quantifier: it can neither be followed by a topic (1b), nor can it follow a narrow focus (1c), at least not within the preverbal sentence zone. [2,3]

---

[1] Two separate papers have been published in Hungarian on *is* and *szintén*: Huszics (2020) and Futó (2020). The present paper integrates the results of these two studies as a basis, and it has the distinguished aim to disseminate their findings in broader international circles.

[2] The grammaticality judgments (on a six-degree scale) referred to throughout the paper are based on the – sufficiently uniform – mother-tongue intuition of the six authors. Nevertheless, there might be speaker-dependent and/or dialectal differences; these require future research using statistical methods. The following abbreviations are used in the glosses:

　　(i)　　case suffixes: ACC(USATIVE), DAT(IVE), INS(TRUMENTALIS), SUP(ERESSIVE);
　　(ii)　agreement suffixes: 1SG ... 3PL; 2OBJ (the object is in the second person);
　　(iii)　other suffixes on nouns: PL(URAL), ADJ(ECTIVALIZER);
　　(iv)　affixes on verbs: PREV (preverb), COND(ITIONAL), INF(INITIVALIZER).

[3] Hungarian is a discourse-configurational language with a [Top ... Top Q ... Q Foc] preverbal operator zone on the left periphery of sentences. The order of these topics, quantifiers and (narrow) foci in this zone corresponds to their scope hierarchy. In Hungarian, a verb is often accompanied by a preverb, immediately preceding the

(1) a.   *Péter **is** felemelte   a   zongorát.*
          Péter   also   preV.lifted.3SG   the   piano.ACC
          'Péter also lifted the piano.'
     b.   **Péter **is** a   zongorát   felemelte.*
          Péter   also   the   piano.ACC   preV.lifted.3SG
     c.   **Csak a   zongorát Péter **is** fel-emelte   / [emelte   fel].*
          only   the   piano.ACC Péter also   preV-lifted.3SG / lifted.3SG   preV

Is is known to inherently have distributive meaning (É. Kiss 2002: 110). That is, (1a) means that there has to be at least one other person, apart from Péter, who lifted the piano, and also, that the sentence cannot be assigned the interpretation that 'Péter belongs to a group the members of which lifted the piano together'.[4]

As shown in (2), *szintén* is also associated with distributive meaning; that is, (2a) also means that there has to be another person besides Péter to have lifted the piano. It also holds true that, similarly to *is*, in the preverbal sentence zone a *szintén*-expression can neither be followed by a topic (2b) nor can it follow a narrow focus (2c).

(2) a.   *Péter **szintén** felemelte   a   zongorát.*
          Péter   also   preV.lifted.3SG   the   piano.ACC
          'Péter also lifted the piano.'
     b.   **Péter **szintén** a   zongorát   felemelte.*
          Péter   also   the   piano.ACC   preV.lifted.3SG
     c.   **Csak a   zongorát Péter **szintén** fel-emelte   / [emelte   fel].*
          only   the   piano.ACC Péter also   preV-lifted.3SG / lifted.3SG   preV

2.2. More than one **is** and/or **szintén** *within one clause*

The second syntactic property that the two particles share is that there cannot be two (or more) *is-* or *szintén*-occurrences in the preverbal zone (3a-a'); this was pointed out by Brody (1990) in connection with *is*.

(3) a.   **Péter **is** a   zongorát **is** felemelte.*
          Péter   also   the   piano.ACC also   preV.lifted.3SG
          Intended meaning: 'Péter also lifted the piano, too.'
     a'. **Péter **szintén** a   zongorát **szintén** felemelte.*
          Péter   also   the   piano.ACC also   preV.lifted.3SG

---

verb stem *by default*. Narrow focus, however, triggers a special order where the verb stem precedes the preverb. For a detailed analysis on Hungarian word order, see É. Kiss (2002).

[4]   *Is* might also have a collective reading but only in extremely peripheral/forced contexts such as the following progressive construction:
  (i)   *Péter is   emelte   (?)('felfelé /??fel)   a   zongorát.*
        Péter   also   lifted.3SG   (upwards / preV$_{up}$) the piano.ACC
        'Péter was also taking part in lifting the piano.'
      However, due to the incomplete event-structures of these constructions, it is problematic to interpret the collective/distributive difference. Potential collective readings of *is* do not form the topic of this paper.

    b.   *Péter* **is** *felemelte   a   zongorát* **is**.
           Péter also preV.lifted.3SG the piano.ACC also
           'Péter also lifted the piano, too.'

    b'.   *Péter* **szintén** *felemelte   a   zongorát* *****szintén**/ˇ**is**.
           Péter also   preV.lifted.3SG the piano.ACC also/also

    c.   (?)*Péter* **szintén** *a   zongorát* **is** *felemelte.*[5]
           Péter   also   the piano.ACC also preV.lifted.3SG
           'Péter also lifted the piano, too.'

    c'.   *ᵗ?Péter* **is** *a   zongorát* **szintén** *felemelte.*
           Péter   also the piano.ACC also   preV.lifted.3SG
           Intended meaning: 'Péter also lifted the piano, too.'

    c''.   *Jóska* **minden** *cikket   elolvasott,   és* **minden** *kötelező   cikket*
           Jóska   every   paper.ACC preV.read.3SG and every   obligatory paper.ACC
           Péter   (?)**is** / ᵗ?**szintén** *elolvasott.*
           Péter   also / also   preV.read.3SG
           'Jóska has read each paper, and Péter, too, has read all the obligatory papers.'

However, in contrast to fully acceptable sentences that contain a preverbal and a postverbal *is*-phrase (3b), those clauses that contain two *szintén*-expressions are unacceptable (3b'). The sentence only becomes acceptable if the second *szintén* is replaced with an *is* (3b',c). However, this does not work the other way round (3c'). A radical difference between the syntax of *szintén* and *is* is illustrated in (3c''): in contrast to *is*, *szintén* is unacceptable after a quantified expression in the pre-V zone.

2.3. The "affiliation" of **is** and **szintén**

Finally, the two particles also share the property that they can both pertain to the noun phrase that they immediately follow (4) – which they primarily do.

(4)   *Péter Julit   Marinak* **is**   / **szintén** *bemutatta.*
        Péter Juli.ACC Mari.DAT also / also   preV.introduced.3SG
        'Péter introduced Juli to Mari, too. [Péter introduced Juli to some people (e.g. to Ili, Kriszti and Olga), and he also introduced her to Mari.]'

3. Differences between *is* and *szintén*

In what follows, we discuss several differences between *is* and *szintén*, which will lead us to extend our scope of interest to a wide range of further areas.

---

[5]   This judgment concerns the (undoubtedly less obvious) reading where the constituent *[zongorát is]* bears a special kind of stress (and interpretation) – characteristic of quantifiers in the comment/predicate zone, as described by Szeteli and Alberti (2018). That is, the same word order is unacceptable if the constituent in question is stressed as a topic-like *is*-expression. This difference suggests that the position of *szintén* is exactly after the topic zone and before the comment/predicate zone, in the sense as Szeteli and Alberti (2018) use the terms.

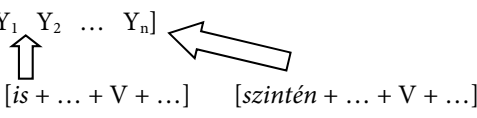### 3.1. *Ordered n-tuples in the scope of 'also'*

Both *is* and *szintén* can pertain to ordered n-tuples of (nominal) constituents, but in different ways, that is, by different word order options. *Szintén* can pertain to noun phrases in a number of ways. It can pertain to the noun phrase immediately preceding it (see meaning 1 in (5)), it can pertain to two noun phrases immediately preceding it (see meaning 2 in (5)), and it can even pertain to several noun phrases in the same way (see meaning 3 in (5), for instance).

(5)   *Péter   Julit   Marinak   **szintén** bemutatta.*
        Péter   Juli.ACC   Mari.DAT   also   preV.introduced.3SG
        meaning 1:   'Péter introduced Juli to Mari, too. [Péter introduced Juli to some people (e.g. to Ili, Kriszti and Olga) and he also introduced her to Mari.]'
        meaning 2:   'Péter introduced Juli to Mari, too. [Péter introduced some people to some people (e.g. Ili to Kriszti and Lajos to Olga) and he also introduced Juli to Mari.]'
        meaning 3:   'Péter introduced Juli to Mari, too. [Some people introduced some people to some people (e.g. Dezső introduced Ili to Kriszti and Ede introduced Lajos to Olga) and Péter introduced Juli to Mari, too.]'

As for *is*, it has to appear immediately after the first element of the ordered tuple, while the other elements of the tuple are preceded by the verb, see (6a-b). (6c) presents the two sentence schemes. (6d) gives a straightforward semantic restriction on the build-up of the tuples in the scope of is: only referential – or more generally, independently interpretable – expressions can function in these tuples. Proper idiom parts (with no independent reference and/or interpretation) cannot appear in these tuples (that is why (6d) has only one idiomatic reading, in contrast to the two literal readings). This restriction also holds for tuples in the scope of *szintén*.

(6)   a.   *Péter **is**   bemutatta   Julit   Marinak.*
               Péter   also   preV.introduced.3SG   Juli.ACC   Mari.DAT
               meaning 1:   'Péter introduced Juli to Mari, too. [Some people introduced Juli to Mari (e.g. Ili, Kriszti and Olga) and Péter also introduced Juli to Mari.]'
               meaning 2:   'Péter introduced Juli to Mari, too. [Some people introduced some people to some people (e.g. Dezső introduced Ili to Kriszti and Ede introduced Lajos to Olga) and Péter introduced Juli to Mari, too.]'[6]
        b.   *Péter Julit   **is**   bemutatta   Marinak.*
               Péter   Juli.ACC   also   preV.introduced.3SG   Mari.DAT
               meaning 1:   'Péter introduced Juli to Mari, too. [Péter introduced some people to Mari (e.g. Betti and Aliz) and Péter also introduced Juli to Mari.]'
               meaning 2:   'Péter introduced Juli to Mari, too. [Péter introduced some people to some people (e.g. Ili to Kriszti and Lajos to Olga) and he also introduced Juli to Mari.]'

---

[6]   The readings are associated with somewhat different stress patterns. Essentially, the rule is that if a phrase belongs to the tuple in the scope of *is* in a particular reading, it bears heavier stress than when it does not belong to the scope of *is*. As a result, there is a (possible but less available) third reading with Péter and Mari in the scope of *is* (marked by heavier stress). The following situation can be presented as an example for this reading: [Some people introduced Juli to some people (e.g. Ili to Dezső, Kriszti to Olga), and Péter also introduced Juli to Mari.].

   c. X_1 X_2 ... X_k $[Y_1$ Y_2 ... $Y_n]$

            $[is + ... + V + ...]$ $[szintén + ... + V + ...]$

   d.   *Péter* ***is*** *megütötte*   *a*   *bokáját.*

        Péter also preV.hurt.3SG the ankle.POSS3SG.ACC

literal meaning 1:	'Péter hurt his ankle, too. [Some people (e.g. Dezső and Jóska) hurt their ankles, and Péter also hurt his ankle.]'
literal meaning 2:	'Péter hurt his ankle, too. [Some people hurt themselves at various places (e.g. Dezső hurt his head, and Jóska hurt his arm), and Péter also hurt his ankle.]'
idiomatic meaning:	'Péter is also going to pay for this.'

3.2. Stress on is *and* szintén

Another difference between the two particles *is* that *szintén* has stress and it can appear as a separate clause, which two facts are presumably interrelated. This behavior of *szintén* can be illustrated with a short conversation set in a restaurant, for instance (see (7a)). *Is* is obligatorily unstressed—it has to cliticize on the constituent that precedes it—and, due to this fact, it cannot appear independently, as shown in (7a).

(7) a.   A: *Vizet*   *kérnék.*

             water.ACC   ask.COND.1SG

             'I would like to have some water.'

       B: ***Szintén***. / **Is*.

             also   / also

             'Me, too.'

  b.   ⟨*Sajnos*⟩   *Péter* ⟨...⟩ *Julit*[#]⟨...⟩ *Marinak* ⟨...⟩ ***szintén*** *bemutatta.*

      unfortunately Péter   Juli.ACC   Mari.DAT   also   preV.introduced.3SG

      The intended meaning is as meaning 2 in (5): 'Péter introduced *Juli* to *Mari*, too. [Péter introduced some people to some people (e.g. Ili to Kriszti and Lajos to Olga) and he also introduced Juli to Mari.]'

      The tuple considered under the intended meaning: [Julit, Marinak]

In contrast to *is*, *szintén* does not cliticize on the last noun phrase in the tuple that it pertains to. Moreover, in a sentence like (5), high sentence-adverbials such as *sajnos* 'unfortunately', *szerencsére* 'fortunately', *szerintem* 'according to me', *állítólag* 'allegedly', *valószínűleg* 'probably' (Surányi 2008) can readily be inserted between the given noun phrase and *szintén* (7b). As also presented in (7b), high adverbials can freely appear in several earlier positions on the left periphery of such sentences, with one restriction: they cannot appear inside the sequence of noun phrases in the scope of *szintén* under the intended meaning (marked by $[Y_1$ Y_2 ... $Y_n]$ in the scheme given in (6c)).

3.3. Is, *and only* is, *as a pragmatic marker*

Is can be replaced with *szintén* only if it functions as a distributive quantifier. If *is* has a pragmatic-marker function in the sense as Aijmer *et al.* (2006) describe it, their replacement results in ungrammatical constructions, as shown in (8). That is, *is* can function as a pragmatic marker while *szintén* cannot. We will return to the topic of the particle functioning as a pragmatic marker later in more detail in Section 5.

(8)   a.   *Kátyúztak   is   / *szintén, meg   nem   is   / *szintén.*
               fixed_road.3PL also   / also   and   not   also / also
               'They both fixed the roads, and they didn't. [They fixed the road in theory, but they didn't do a very good job at it.]'

        b.   *Kit   is   / *szintén láttunk tegnap?*
               who.ACC also   / also   saw.1PL yesterday
               'Again, who did we see yesterday?'

3.4. Szintén, *and only* szintén, *as a ground-preserving discourse marker*

A further difference between *is* and *szintén* is that *is* always pertains to the element which it follows, while *szintén* can also pertain to the constituent following it. In the latter case, *szintén* has a function which can be referred to as a "discourse preserving" function (see Huszics 2020). We try to convey this function in the translations under (9). In this function, *szintén* can be interpreted as a coherence-signaling discourse marker, as proposed by Aijmer *et al.* (2006) in their terminology.

(9)   a.   *Tegnap   elfogták   a   bankrablót.*
               yesterday preV.captured.3PL the bank_robber.ACC
               (És) szintén tegnap   Tóth   bíró   elítélte   a   bűntársait.
               and   also   yesterday Tóth judge preV.sentenced.3SG the accomplice.PL.ACC
               'The bank robber was caught yesterday. (And) also yesterday, judge Tóth sentenced his accomplices.'

        b.   *Ibi hozta   be   a   gyilkost.*
               Ibi brought.3SG preV the murderer.ACC
               (És) szintén Ibi oldotta   meg   a   bankrablási   ügyet.
               and   also   Ibi solved.3SG preV the bank_robbery.ADJ case.ACC
               'It was Ibi who brought in the suspect. (And) it was also Ibi who solved the bank robbery case.'

        b'.   ??[*Ibi be-hozta   a   gyilkost.*
               Ibi   preV-brought.3SG the murderer.ACC
               (És) szintén Ibi meg-oldotta   a   bankrablási   ügyet.]
               and   also   Ibi preV-solved.3SG the bank_robbery.ADJ case.ACC
               'Ibi brought in the suspect. (And) it can also be said about Ibi that she solved the bank robbery case.'
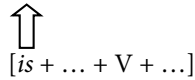
If the ground-preserving *szintén* pertains to an argument (9b-b') and not to an adjunct (9a), the argument has to appear in focus position (Huszics 2020). It requires future research to establish the kinds of satellites that can appear both in topic and in focus position, and the kinds of satellites that can appear only in focus position. At this point though, it can be stated that adjuncts can appear in both positions (presumably due to their independent interpretability, which is expectable with topic-like functions), while central arguments (viz. nominative and accusative case-marked ones) can only appear in focus, with non-central arguments seemingly showing an in-between (but uneasily characterizable) behavior.

4. Syntactic representations

This section provides syntactic representations for various constructions containing *is* or *szintén*.

4.1. The syntactic representation of is

First let us consider the structure belonging to (6b), repeated here as (10) in meaning 2. This meaning can be regarded as a realization of the general scheme presented in (6c), as follows: X_1=Péter, k=1, n=2, [Y_1, Y_2] = [Juli, Mari], see (10b).

(10)   a.   *Péter Julit   is   bemutatta   Marinak.*
       Péter Juli.ACC also   preV.introduced.3SG Mari.DAT
       'Péter introduced Juli to Mari, too. [Péter introduced some people to some people (e.g. Ili to Kriszti and Lajos to Olga) and he also introduced Juli to Mari.]'

   b. X_1 X_2 … X_k [Y_1 ~~Y_2 … Y_n~~]   [~~Y_1~~ Y_2 … Y_n]

                         ⇧
                     [*is* + … + V + …]

The structure is based on Grohmann's (2003) theory of Prolific Domains – in particular, on the version proposed for Hungarian by Farkas and Alberti (2017). The triangle under the vP node signals that the internal build-up of the thematic domain (ΘΔ) is irrelevant here. In the ΦΔ, CentPs host central arguments (i.e. the subject and the object), while NonCentPs host non-central arguments. The order of CentPs and NonCentPs is free; it is based on Behaghel's Law, often used by É. Kiss (e.g. 2009), which predicts that the optimal order of the satellites of V at the right periphery of the Hungarian sentence is the following: heavy phrases tend to come last. In harmony with this principle, the basically free word order (determined mainly by the Behaghel's Law) appearing in Hungarian in the postverbal zone is expressed as such: the highest phrase in the ΦΔ is AspP, whose modifier is occupied by the preverb expressing the aspect of the sentence, and to whose head the verb moves.
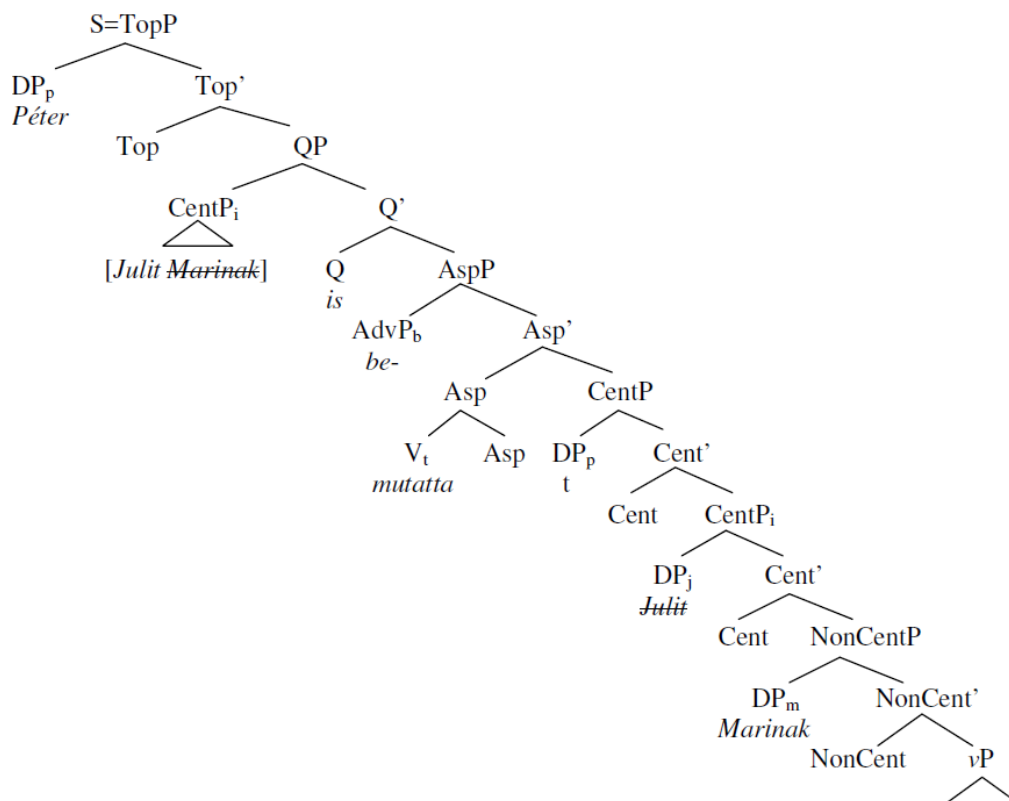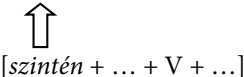

Figure 1: *The syntactic representation of (10a)*

As the particle *is* pertains to a whole proposition in (10), a whole CentP has to appear (in the $\Omega\Delta$) in the modifier of the quantifier, the head of which is occupied by the particle itself. Due to the fact that *is* is obligatorily unstressed and it has to cliticize on the head of the constituent it pertains to, the CentP is only partially spelt out in the modifier in question ([Y_1, ~~Y_2~~]); the rest of the CentP is spelt out in its base-generated position ([~~Y_1~~, Y_2]).[7] The subject (Péter) appears as a topic, since it does not belong to the ordered n-tuple in question (it plays the role of X_1 in (10b) and (6c)).

4.2. The syntactic representation of szintén

Figure 2 presents the syntactic structure of the corresponding sentence containing *szintén* (see (5) in meaning 2), repeated here as (11).

(11) a.   *Péter Julit   Marinak **szintén** bemutatta.*
         Péter Juli.ACC Mari.DAT also   preV.introduced.3SG
         'Péter introduced Juli to Mari, too. [Péter introduced some people to some people (e.g. Ili to Kriszti and Lajos to Olga) and he also introduced Juli to Mari.]'

   b. X_1 X_2 … X_k [Y_1 Y_2 … Y_n]   ~~[Y_1 Y_2 … Y_n]~~

                                      ⇧

                              [*szintén* + … + V + …]

This meaning can also be regarded as a realization of the general scheme presented in (6c), with the same specification: X_1=Péter, k=1, n=2, [Y_1, Y_2] = [Juli, Mari], see (11b). The crucial difference is that *szintén* has no "addiction" to insertion but behaves as a Cinquean middle (predicate-) adverbial, as proposed by Surányi (2008) (on the basis of word-order facts and the stressed status of *szintén*), which projects as an AlsoP left-adjacent to the field of the complex predicate (AspP here). Witnessed by (7b) in 2.2, projections of high sentences-adverbials (such as sajnos 'unfortunately') can appear between the lower TopP and this AlsoP (apart from when they appear between the two TopPs and above the higher TopP). That is why the specifier of AlsoP is assumed to only contain the trace of the remnant with the phonetic realization Julit Marinak. This latter sequence is considered to form a unit, in order to account for the fact that no high adverbial can appear between Julit and Marinak (as also illustrated in (7b)). This unit is assumed to occupy the specifier of a topic, instead of that of a quantifier, due to the data shown in (3c') in 1.2: this way, the ill-formed status of the word order in question can be attributed to the rule that a quantifier cannot take scope over a topic.[8] This implies the following complex semantics for *szintén*: its use relies on a complex predicate in an earlier clause, and

---

[7]   The *partial spell-out* technique was first applied to Hungarian by Surányi (2009).

[8]   The same rule accounts for the well-formed status of (3c): the noun phrase before the AlsoP is a Topic, which is permitted to take scope over the *is*-quantifier after the AlsoP. AlsoP is the leftmost element of the comment / predicate zone of the Hungarian sentence structure, immediately after the topic zone. Quantifiers can appear in both sentence zones (Szeteli and Alberti 2018), resulting in this general structure (H: high adverbials, *: Kleene star, which means that 0, 1, 2 or more copies of the given category can occur):

[Topic zone H* Top* H* Top* H* Q* H*] [Comment zone Also Q* Foc …]

szintén associates the new copy of this (in the given clause) with a new tuple of referents.[9] The distributive character can be explained as follows: the new topic does not refer to a participant in the original (or old) eventuality itself, but to a participant in an eventuality which is of the same type as the antecedent eventuality, but another instance of that type.
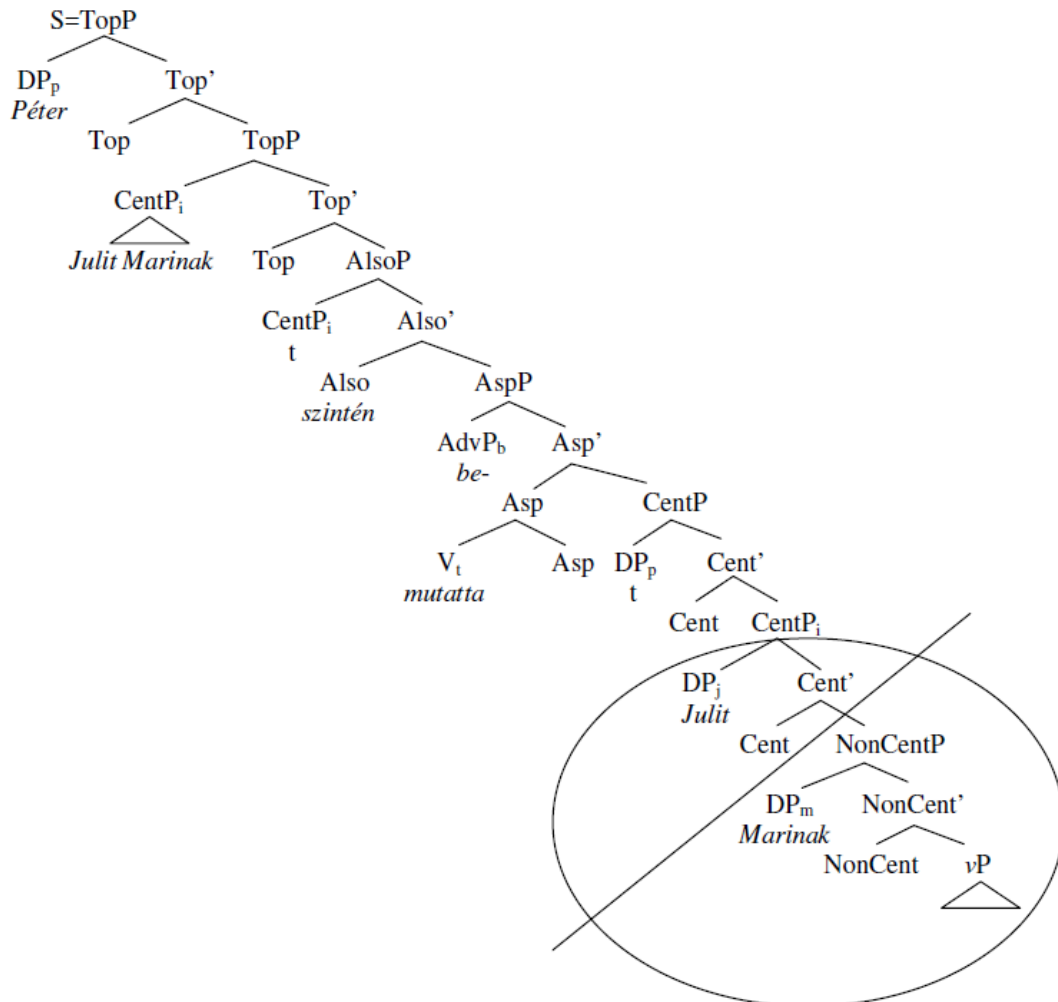

Figure 2: *The syntactic representation of (11a)*

4.3. *The syntactic representation of the "ground preserving"* szintén

The final structure we consider is the one where *szintén* has a discourse preserving function; see (9a), for instance, repeated here as (12)). Just like in 3.2, its semantics relies on the association of something old with something new; except here, certain elements in the topic zone of an earlier clause are exploited. This anchoring information is associated with a new complex predicate, whose projection appears in the complement of the given Also head.

(12)   (*És*)   ***szintén***   *tegnap*   *Tóth*   *bíró*
        and   also   yesterday Tóth   judge

---

[9]   The new copy of the antecedent predicate can be elided. *Julit Marinak szintén* – this verbless sequence of words, for instance, can serve as a well-formed alternative to (11a) in contexts with appropriate antecedents.

elítélte   *a*   *bűntársait.*
preV.sentenced.3SG   the   accomplice.POSS3SG.PL.ACC
'[The bank robber was caught yesterday.] (And) also yesterday, judge Tóth sentenced his accomplices.'

As formulated in Huszics (2020), the specifier of the AlsoP contains the antecedent clause, which is then to be regarded as still "active" in the discourse: the antecedent clause 'yesterday, the bank robber was caught' is referred to by the anaphoric element pro-Top'.[10] Given the fact that this kind of *szintén* must be cliticized on the constituent in the specifier from the left, the word order changes in the Phonological Form (resulting in the sequence *szintén tegnap*). Here *szintén* functions as a *high* sentence-adverbial (old topic(s) with a new predicate), in contrast to the predicate-adverbial variant of *szintén* (new topic(s) with an old predicate).
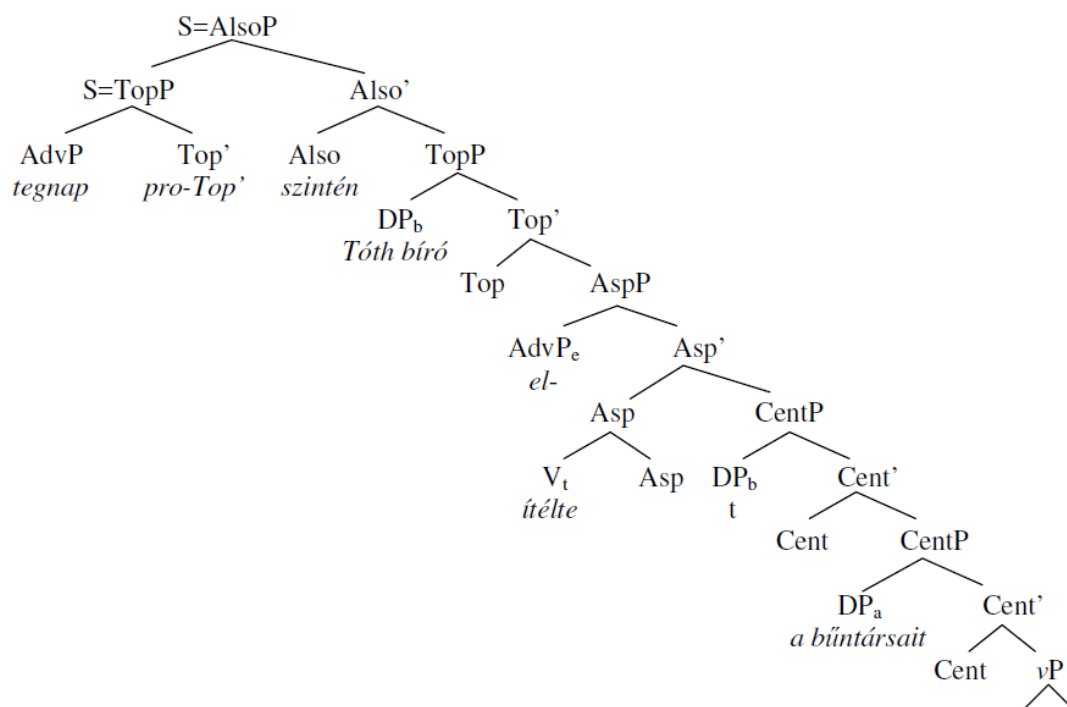

Figure 3: The syntactic representation of (12)

5. *Is* as a pragmatic marker

As mentioned previously (see (8) in Section 3), the particle *is* does not only have a logical role, but it can also function as a pragmatic marker – as used in the terminology proposed by Aijmer *et al.* (2006). In this role, it cannot be replaced with *szintén* (NB: the function of *is* can be called 'logical' if the given function can be expressed by a corresponding *szintén*-expression). In this section, we present how the different kinds of *is* can be systematically differentiated from one another, and how they can be accurately characterized in a formal representational dynamic

---

[10] Bánréti (1992: 789) accounts for various elliptical constructions by using empty proposition-level pronominal expressions. In his example (136'), a pro-V' refers to a situation with a boy looking at a girl. The sentence in question consists of a single word: *Hosszasan* 'at length'. In the same way, *Tegnap* 'yesterday' can serve as a short answer to the question: *When was the robber caught?*

theory of interpretation called ℜeALIS (*Reciprocal and Lifelong Interpretation System*).[11] These observations are the first steps towards revealing the complex polysemic network of the particle *is* in full (Fischer 2006: 13–14), with special regard to the subtle differentiation between pragmatic markers and discourse markers, such as *szintén*. As mentioned in the Introduction, a pragmatic marker in the terminology of Aijmer *et al.* (2006) is a word or expression that does not contribute to the propositional, truth-functional content of an utterance, while a discourse marker signals coherence relations.

5.1. *The pragmasemantic characterization of the emotive* **is**

The example below (13a) serves to illustrate one of those cases where the role of *is* can be "non-logical". Here, *hárman is* (three+also) does not refer to a set of more than three (that is, at least four) students. Instead, *is* yields an exclamative sentence, with the intensional characteristics given in (13b).

(13) a.   *Hárman* **is**   *megoldották*   *ezt*   *a*   *feladatot!*
           three   also   preV.solved.3PL   this.ACC   the   task.ACC
           'There were (as many as) three people who solved this task (separately).' = *e*

      b.   iB↓ ∈ '−3'   [At an earlier point of time, I considered it quite unlikely that *e* should take place.]
           iD ∈ '5   [*e* taking place has a considerable emotional impact on me, either a positive or a negative one]

The pieces of intensional information involved in this situation are presented with the formalism of ℜeALIS. Essentially, the strings (e.g. iBuB↓) represent possible worlds of their holder, which is indicated by the first symbol in them (i=I (the speaker), u=you (the listener), o/o'/o"/… = 'underspecified holders'). The second symbol represents the modality of the possible world that the holder of the world (indicated by the first symbol) associates with the propositional content of the sentence. It can be, among others, a belief (B), a desire (D) or an intension (I). The modality is followed by an element of the set T={↓,↑,•}, which indicates the temporal properties of the world ('↓ ' for past, '↑' for future, '•' for present, in the string usually unmarked). If such a string is followed by a new holder, a new modality and a new temporal property, it means that it is embedded into the previous wordlet. This technique of reciprocity (cf. ℜeALIS) makes it possible to handle, for instance, the thoughts that interlocutors attribute to each other in the course of mind reading. A scale is also introduced for the values referring to false (−5), true (+5), and underspecified (0) information, and to various degrees between them (which more or less correspond to Farkas and Roelofsen's (2017) credence levels). As for the apostrophes, relative to the value −3, the complex symbol '−3' denotes the interval {−5, −4,

---

−3, −2, −1} around the precise value −3. The relation iBuB↓ ∈ '+4·−5 thus means that "I am more or less convinced that you thought that a given information unit *e* is false."

Thus, (13b) provides possible-world indices. The first index defines the possible world of certain earlier assumptions of the addresser (modified later).[12] The second one "collects" states of affairs immensely desirable or, on the contrary, definitely undesirable for the addresser. We claim that this underspecified situation should be regarded as a linguistically and conventionally encoded part of the meaning of the given type of *is* (cf. Farkas and Roelofsen 2017: 237). In a particular conversation, the polarity value can typically be decided on the basis of world knowledge. Success in solving tasks is typically desirable, but the particular context may imply the opposite polarity value.

5.2. *The pragmasemantic characterization of* is *meaning 'and so it happened'*

The second type of *is* we discuss is the one that can be referred to as the *Úgy is lett* 'And so it happened' type. This type is mentioned in Alberti and Farkas (2018: 41), where its syntactic properties are accounted for. Here (see (14)), we present the crucial elements of the pragmasemantic contribution of the type to the discourse, based on the three examples in (14a-a").

(14)   a.   *Gondoltuk,   hogy   havazni   fog.   És   havazott   is!*
            thought.1PL   that   snow.INF   will.3SG   and   snowed.3SG   also
            'We thought that it would snow. And so it did.'

       a'.   *Rossznak   tűnik.*   [a bit later]   *És   rossz   is!*
             bad.DAT   seem.3SG   and   bad   also
             'It looks bad. *[a bit later]* And so it is.'

       a".   *Marinak   akartalak   bemutatni.*
             Mari.DAT   wanted.2OBJ.1SG   preV.introduce.INF
             És   Marinak   is   mutattalak   be!
             and   Mari.DAT   also   introduced.2OBJ.1SG   preV
             'I wanted to introduce you to Mari specifically. And so I did.'[13]

       b.   iB↓ ∈ '+3'   [I considered it likely that *e* should take place]
            iD ∈ '5   [*e* taking place has a considerable emotional impact on me, either a positive or a negative one]

5.3. *The pragmasemantic characterization of the* is *of refutation*

Compared to the intensional profile in (13b), the one in (14b) also contains a substantial emotional impact, but here, the anticipatory epistemic state is positive. In example (14a"), this is not expressed explicitly in the sentence before the one including the *is*, but it is obvious that committed volition tends to yield success.

---

[12] In ℜeALIS, the term *addresser* is used with the aim of distinguishing (and then truth-conditionally comparing) the 'addresser role' as a conventionalized intensional profile (to be calculated based on sentence types and discourse markers) from 'speaker' – as a conscious psychological ego with permanently changing information states.

[13] The first syllable of (14a") in both occurrences of the noun *Marinak* is pronounced with the (same) peculiar narrow-focus stress.

The third type we discuss is a kind of denial, as presented in (15b), an answer to (15a). This construction is typically used to explicitly refute some presupposition within the content of the antecedent statement. Hence, a detailed formal analysis of this type requires the "lifelong" character of ℜeALIS (cf. ℜeA<u>L</u>IS). This means that the basis of interpretation contains the interlocutors' momentary information states as lifelong-built (Alberti 2000) discourse representation structures (Kamp *et al.* 2011). These structures can obviously contain both the contextual information in the interlocutors' episodic memories and the scripts in their semantic memories (Leiss 2014), required to calculate the connection between the (b-b')-sentences and the (a)-sentence. The (15a+b') dialogue, for instance, can be explained with reference to a script which is available to the interlocutors and which contains information on the relation between dancing and potential dance lessons.

(15) a. A: *Tegnap   táncoltam   Ilivel   a   bulin.*
                  yesterday   danced.1SG   Ili.INS   the   party.SUP
                  'I danced with Ili yesterday at the party.'

      b. B: *Nem   **is**   Ili   volt   a   lány   neve.*
                  not   also   Ili   was   the   girl   name.3SG
                  / *Nem **is**   tegnap   volt a   buli.*
                  / not   also   yesterday was   the   party
                  '(But) the girl's name was not (even) Ili. / (But) the party was not (even) yesterday.'

      b'. B: *Nem **is**   tudsz   táncolni.*
                  not   also   can.2SG   dance.INF
                  'You cannot even dance.'

      c. iB ∈ −5   [I think that *e* is false]
          iB<u>uB</u> ∈ −5   [I think that you are aware of the fact that *e* is false]
          iBuIi<u>B</u>↑ ∈ +5   [it seems to me that you intend to have me believe that *e* is true]
          <u>iD</u>uIiB↑ ∈ '−5   [I wish you intended to tell me the truth, according to which *e* is false]

(15c) presents the possible-world indices that we propose as the definitive factors of the conventionalized intensional profile of the "refutative" *is*-construction. The information in question (denoted by *e* above) is the content of the sentence in (15a). Essentially, the profile expresses the addressee's stance according to which the addresser should stop pretending as if (15a) were true.

5.4. *The pragmasemantic characterization of the* **is** *of "contradictions"*

The fourth type of *is* we discuss displays the form [p and not p] of a logical contradiction, which is regarded as a non-contingent but constantly false expression in logics (16a-a'). The given double-*is*-construction, however, does provide some contingent intensional contribution. It becomes possible like this: the two clauses that the construction consists of concern two different possible worlds; and it is not impossible at all but definitely typical that alternative possible worlds support a proposition and also its opposite.

(16) a.   *Kátyúztak*   **is**, *meg* *nem* **is.**
           fix_road.PAST.3PL also and not also
           'They both fixed the roads, and they didn't.'
           [They fixed the road in theory, but they didn't do a very good job at it.]'
           http://www.vasarhely24.com/vasarhely/katyuztak-is-meg-nem-is-a-rarosin (20200109)

      a'. Burgonya *is*,   *meg* *nem* *is.*
           potato   also and not also
           '[Sweet potatoes] are both potatoes and not.'
           [Certain properties of potatoes, in contrast to other properties, hold for sweet potatoes.]'
           https://www.magyarszo.rs/hu/3826/hetvege/192403/Burgonya-is-meg-nem-is.htm (20200109)

      b.   iBo'B ∈ +5   [I believe that some people think that *e* has taken place]
           iBo"B ∈ −5   [I consider that some people think that *e* has not taken place]

      b'. iB' ∈ +5   [a set *e'* of criteria (characterizing potatoes) holds true (for sweet potatoes)]
           iB" ∈ −5   [a set *e"* of criteria (characterizing potatoes) does not hold true (for sweet potatoes)]

Based on the analyses of a wide range of relevant Internet data, we hypothesize that the intensional profile of the construction under investigation is highly underspecified. The common factor in the data is that the interpretation of the same content in the same possible world is somehow avoided. Apart from this, the interpretation of the utterance is often extremely vague. Regarding the sentence in (16a), the profile proposed in (16b) accounts for the apparent contradiction in that it assumes that the addresser articulates the opinions of different groups (o' and o") of people. As for the profile (16b'), proposed for the sentence in (16a'), our analysis relies on the script- and schema-based technique of interpretation in ℜeALIS, as exemplified in (15b'). For something to qualify as a potato requires the satisfaction of a set of criteria more or less known to the interlocutors. Thus, the two clauses in (16a') should be regarded as ones that concern two different sets of properties related to sweet potatoes (viz. the potato-like and the non-potato-like properties, respectively).

5.5. *The pragmasemantic characterization of the basic metacognitive* is

The last type of *is* we discuss here to demonstrate the non-logical uses of *is*—where it cannot be replaced with *szintén*—is the one that can be referred to as the metacognitive *is*; the nearest English equivalent for this *is* could be 'again' (17a).

(17) a.   *Kit*   **is**   *láttunk*   *ott*   *tavaly?*
           who.ACC also see.PAST.1PL there last_year
           'Again, who did we see there last year?'

      b.   (i)   iB ∈ '0'   [I do not know whether *e* is true or false]
           (i')   iB*u*B ∈ +5   [I think that you know the truth]
           (i")   iIuI↑*u*B↑ ∈ +5   [I intend to have you prepared to tell me the truth]
           (ii)   iB↓ ∈ +5   [there was a time when I knew the truth]
           (ii')   iBuBi*B↓* ∈ +5   [I think that you are aware of the fact that there was a time when I knew the truth]

The intensional profile of the metacognitive *is* (at least, of this particular subtype, see Futó (2020)) is presented in (17b.ii-ii'). The addresser is scanning his/her own lifelong mind representation and realizes that there is some information that he/she had had earlier (17b.ii)

but is now missing from his/her current information state (17b.i). They also realize that, luckily, this is a piece of information that they are likely to (have) share(d) with the addressee (17b.ii'). Note that the profile elements given in (17b.i-i") are necessary for elucidating the functioning of the *wh*-word+*is* construction (the addresser asks for the missing information), but they belong to the general interrogative intensional profile, discussed in Alberti *et al.* (2019b) and Viszket *et al.* (2019), among others.

6. Summary

The purpose of the paper was to discuss two Hungarian particles with the same logical core meaning ALSO: *is* and *szintén*. Since *is* has been extensively studied in Hungarian generative literature (e.g. Brody 1990 and É. Kiss 1992, 2002) while *szintén* has scarcely been paid attention, we described the two particles in relation to each other, using the already available descriptors for *is* as a point of departure for the description of *szintén*. The parallel analysis of the two particles yielded novel observations for the syntax and pragmasemantics of both particles, as summarized below (see Section 2 for details):

I.   Both *szintén* and *is* have a distributive meaning.

II. Neither *is*, nor *szintén* can be followed by a topic, and neither of them can follow a (narrow) focus.

III. *Is* and *szintén* can be combined within one clause, with certain restrictions on word order.

IV. Both *szintén* and *is* can pertain to ordered n-tuples of noun phrases (as a generalization of the case n=1), but in syntactically different ways.

V.   *Is* is obligatorily unstressed, while *szintén* bears stress, with the exception of its ground-preserving use, when it can be regarded to function as a discourse marker (Aijmer *et al.* 2006).

VI. *Szintén* can appear as a separate clause, while *is* cannot.

VII. *Is* can function as a pragmatic marker (Aijmer *et al.* 2006) while *szintén* cannot.

After an overview of the similarities and differences (Sections 2 and 3), we demonstrated how the syntactic characteristics of the two particles can be captured by the *partial spell-out* technique used in some branches of minimalism (Surányi 2009), and we classified *szintén* as a high sentence-adverbial (Surányi 2008) – in contrast to the clitic-like *is* (Section 4). In what followed, we drew up a pragmasemantic system for those usages of *is* where the particle cannot be replaced with *szintén* – i.e. the cases where *is* functions as an intensional/metacognitive pragmatic marker, as opposed to *szintén*, which acts as a coherence-signaling discourse marker (Section 5). For this analysis, we used the framework of the formal representational dynamic theory of interpretation called ℜeALIS (see, for instance, Alberti *et al.* 2016, 2019; Szeteli *et al.* 2019, and Viszket *et al.* 2019).

References

Aijmer, K., A. Foolen, and A.-M. Simon-Vandenbergen. 2006. Pragmatic markers in translation: A methodological proposal. *Approaches to Discourse Particles* 1: 101–114.

Alberti, G. 2000. Lifelong discourse representation structures. *Gothenburg Papers in Computational Linguistics* 00–5: 13–20.

Alberti, G. 2011. *ℜeALIS. Interpretálók a világban, világok az interpretálóban* [Interpreters in the World, Worlds in the Interpreter]. Budapest, Hungary: Akadémiai Kiadó.

Alberti, G., M. Dóla, E. Kárpáti, J. Kleiber, A. Szeteli, and A. Viszket. 2019a. Towards a cognitively viable linguistic representation. *Argumentum* 15: 62–80.

Alberti, G., M. Dóla, E. Kárpáti, J. Kleiber, A. Szeteli, and A. Viszket. 2019b. Kérdéses világaink [Our Worlds in Question(s)]. *Jelentés és Nyelvhasználat* [Meaning and Language Use] 6: 11–35.

Alberti, G., and J. Farkas. 2018. A syntactic solution to the inverse-scope puzzle and the contrastive VP-focus construction in Hungarian based on extraction. In A. Bloch-Rozmej and A. Bondaruk (eds.), *Studies in Formal Linguistics*, 35–54. Berlin, Germany: Peter Lang.

Alberti, G., and J. Kleiber. 2015. ℜeALIS: Discourse representation with a radically new ontology. In L. Veselovská and M. Janebová (eds.), *Complex visibles out there: Olomouc modern language series* 4, 513–528. Olomouc, Czech Republic: Palacký University.

Alberti, G., J. Kleiber, Zs. Schnell, and V. Szabó. 2016. Intensional profiles and different kinds of human minds: "Case studies" about Hungarian imperative-like sentence types. *Linguistics Beyond And Within* 2: 6–26.

Andor, J. 2011. Reflections on Speech Act Theory: An interview with John R. Searle. *International Review of Pragmatics* 3: 1–12.

Bánréti, Z. 1992. A mellérendelés [Coordination]. In F. Kiefer (ed.), *Strukturális magyar nyelvtan* 1 [A Structural Grammar of Hungarian 1], 715–796. Budapest, Hungary: Akadémiai Kiadó.

Brody, M. 1990. Remarks on the order of elements on the Hungarian Focus Field. In I. Kenesei (ed.), *Approaches to Hungarian* 3, 95–121. Szeged, Hungary: JATE.

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], 79–177. Budapest, Hungary: Akadémiai Kiadó.

É. Kiss, K. 2002. *The syntax of Hungarian*. Cambridge, UK: Cambridge University Press.

É. Kiss, K. 2009. Is free postverbal order in Hungarian a syntactic or a PF phenomenon? In N. Erteschik-Shir and L. Rochman (eds.), *The sound pattern of syntax*, 53–71. Oxford, UK and New York, USA: Oxford University Press.

Farkas, D. and F. Roelofsen. 2017. Division of Labor in the Interpretation of Declaratives and Interrogatives. *Journal of Semantics* 34: 237–289.

Farkas, J. and G. Alberti. 2017. The Hungarian HATNÉK-noun expression: A hybrid construction. In A. Bloch-Rozmej and A. Bondaruk (eds.), *Constraints on structure and derivation in syntax, phonology and morphology*, 71–100. Frankfurt am Main, Germany: Peter Lang.

Fischer, K. 2006. Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume. In K. Fisher (ed.), *Approaches to discourse particles*, 1–20. Amsterdam, The Netherlands: Elsevier.

Futó, B. 2020. A logikai *is*-től a pragmatikai *is*-en át a metakognitív *is*-ig [From the logical *is* through the pragmatic *is* to the metacognitive *is*]. In G. Böhm, D. Czeferner, and T. Fedeles (eds.), *Specimina Operum Iuvenum 6*, 13–49. Pécs, Hungary: PTE BTK Scientific Students Circle.

Grohmann, K. K. 2003. *Prolific domains: On the anti-locality of movement dependencies. Linguistik Aktuell 66.* Amsterdam, The Netherlands: John Benjamins.

Huszics, A. 2020. *Szintén* [Also]. In G. Böhm, D. Czeferner, and T. Fedeles (eds.), *Specimina Operum Iuvenum 6*, 50–84. Pécs, Hungary: PTE BTK Scientific Students Circle.

Kamp, H., J. van Genabith, and U. Reyle. 2011. Discourse representation theory. In D. Gabbay and F. Guenthner (eds.), *Handbook of philosophical logic* 15, 125–394. Berlin: Springer.

Kleiber, J. and G. Alberti. 2017. Compositional analysis of interrogative imperatives in Hungarian. *Linguistics Beyond And Within* 3: 117–135.

Lauer, S. 2013. *Towards a dynamic pragmatics.* Ph.D. diss. Stanford University.

Leiss, E. 2014. Modes of modality in an Un-Cartesian framework. In S. Cantarini, W. Abraham, and E. Leiss (eds.) *Certainty-uncertainty – and the attitudinal space in between. Studies in language companion series* 165, 47–62. Amsterdam, The Netherlands: Benjamins.

Surányi, B. 2008. Határozóosztályok és mondattartományok [Adverbial classes and sentence domains]. *Nyelvtudományi Közlemények* [Issues in linguistics] 105: 163–192.

Surányi, B. 2009. "Incorporated" locative adverbials in Hungarian. In K. É. Kiss (ed.), *Adverbs and adverbial adjuncts at the interfaces*, 39–74. Berlin, Germany: Mouton de Gruyter.

Szeteli, A. and Alberti, G. 2018. The interaction between relevant-set based operators and a topic-predicate dimension. *Linguistics Beyond And Within* 4: 161–172.

Szeteli, A., M. Dóla, and G. Alberti. 2019. Pragmasemantic analysis of the Hungarian inferential-evidential expression 'szerint'. *Studies in Polish Linguistics* 14: 207–225.

Viszket, A., A.Hoss, E. Kárpáti, and G. Alberti. 2019. Recalculating: The atlas of pragmatic parameters of developmental disorders. *Linguistics Beyond And Within* 5, 148–160.

Focus-feature and *wh*-feature in the light of pied-piping behavior in Hungarian

Júlia Keresztes

University of Debrecen, Hungary

Abstract

This paper presents the findings of two experiments on pied-piping by a prenominal adjunct in Hungarian focus- and *wh*-constructions. According to Webelhuth (1992) and Cable (2010), pied-piping from adjunct islands is prohibited. The results of the experiments, however, suggest that not only is it possible in Hungarian for the pied-piper to be embedded inside an adjunct island but the embedding of the pied-piper and the movement itself are unrestricted in focus-construction. Pied-piping shows a similar picture in *wh*-constructions with the restriction that the *wh*-element has to be discourse-linked for pied-piping to be allowed.

Keywords: pied-piping, adjunct-islands, focus-feature, *wh*-feature, Hungarian

1. Introduction

This paper presents the findings of two experiments on pied-piping by a prenominal adjunct in Hungarian focus- and *wh*-constructions. Pied-piping is a syntactic operation in which a phrase triggers movement of a larger phrase containing it. This movement is considered to be a rescue strategy in most languages, when movement of the word/phrase is not allowed for an independent grammatical reason, dragging along the phrase containing it can rescue the structure from being ungrammatical. Pied-piping is generally considered acceptable in Hungarian as in (1) and in English as well (2).

(1)   [*Milyen   színű   szoknyát*]   *vettél*   *az   ünnepségre?*
        Which   color   skirt-ACC   bought   the celebration.for
        'What color skirt did you buy for the celebration?'

(2)   a.   [*Which book*] *did you buy in the store?*
        b.   [*Whose brother's book*] *did you take by mistake?*

The motivation for the experiments was provided by data in previous literature that presented pied-piping and the lack thereof as an empirical argument against the existence of a syntactic focus-feature (Horváth 1997, 2010).

(3)   a.   * *az ital,   amit   követelő   vendégektől fél   a   pincer   t*
            the   drink   which-ACC   demanding   guests   fear-3SG   the waiter
            'the drink customers demanding which the waiter is afraid of…'
      b.   * *Mit   követelő   vendégektől   fél   a   pincer?*
            what-ACC   demanding   guests   fear-3SG   the waiter
            'Customers demanding what is the waiter afraid of?'
      c.   *BARACKPÁLINKÁT   követelő   vendégektől   fél   a   pincer.*
            apricot-brandy-ACC   demanding   guests   fear-3SG   the waiter
            'It is customers demanding APPRICOT BRANDY that the waiter is afraid of.'

The structural similarities between *wh-* and focus-constructions prompted the investigation of pied-piping in both constructions. The research questions of the experiments were the following:

Q1:   Is pied-piping permissible in *wh*-constructions and focus-constructions?
Q2:   Is discourse-linking a factor in the availability of pied-piping?

In this paper I present two experiments on pied-piping in Hungarian. First, I will introduce two important accounts of pied-piping; Webelhuth (1992) attributing the availability of pied-piping to a structural position, and Cable (2010) attributing the availability of pied-piping to the existence of Agree between a Q-operator and a Q-feature on the *wh*-phrase. Then I will turn to the theoretical background of Hungarian focus and *wh*-constructions (section 3). Then I will briefly introduce discourse-linking and the motivation for including discourse-linking as a factor in the experiments (section 4). After that, I will present the experiments and their results (section 5), and give a tentative proposal based on the results of experiments (section 6). Finally, I will conclude (section 7).

2. Two accounts of pied-piping

Webelhuth (1992) discusses pied-piping in Germanic languages. His theory of pied-piping involves feature percolation and theta theory: features can percolate from certain positions to the maximal projection and this enables the constituent to undergo pied-piping. He distinguishes the positions in a given phrase by their ability to act as a pied-piper of the phrase: specifiers and heads are pied-pipers, while complements and adjuncts are not. Webelhuth claims that it is not only feature percolation that counts. The theta-criterion dictates that theta marked arguments can only be in the derivation once – as the theta-criterion demands that a theta-marked position must be a part of a chain containing exactly one argument. Theta-marked positions in his theory are exactly the positions from which constituents cannot undergo pied-piping. This means that complements and modifiers/adjuncts are theta-marked and therefore cannot be pied-pipers of a given phrase. Whereas specifiers and heads of a phrase are not theta-marked, they are pied-pipers for the phrase.

(4)   a.   ** I wonder [give a talk where] John will t.*

        b.   ** I wonder [the party where] John will enjoy t.*

(5)   a.   *[To whom] did John talk?*

        b.
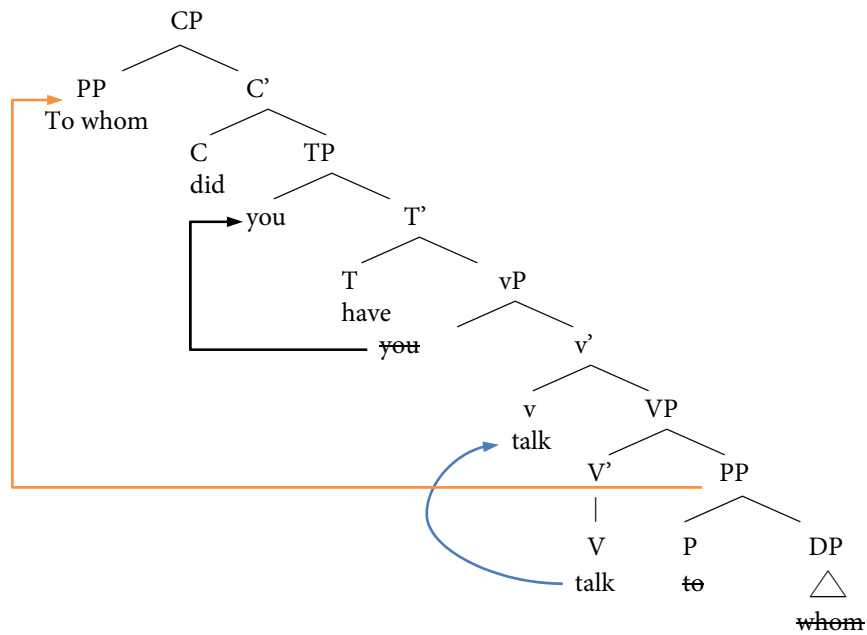
        c.   *I wonder [whose mother] you have seen.*

        d.
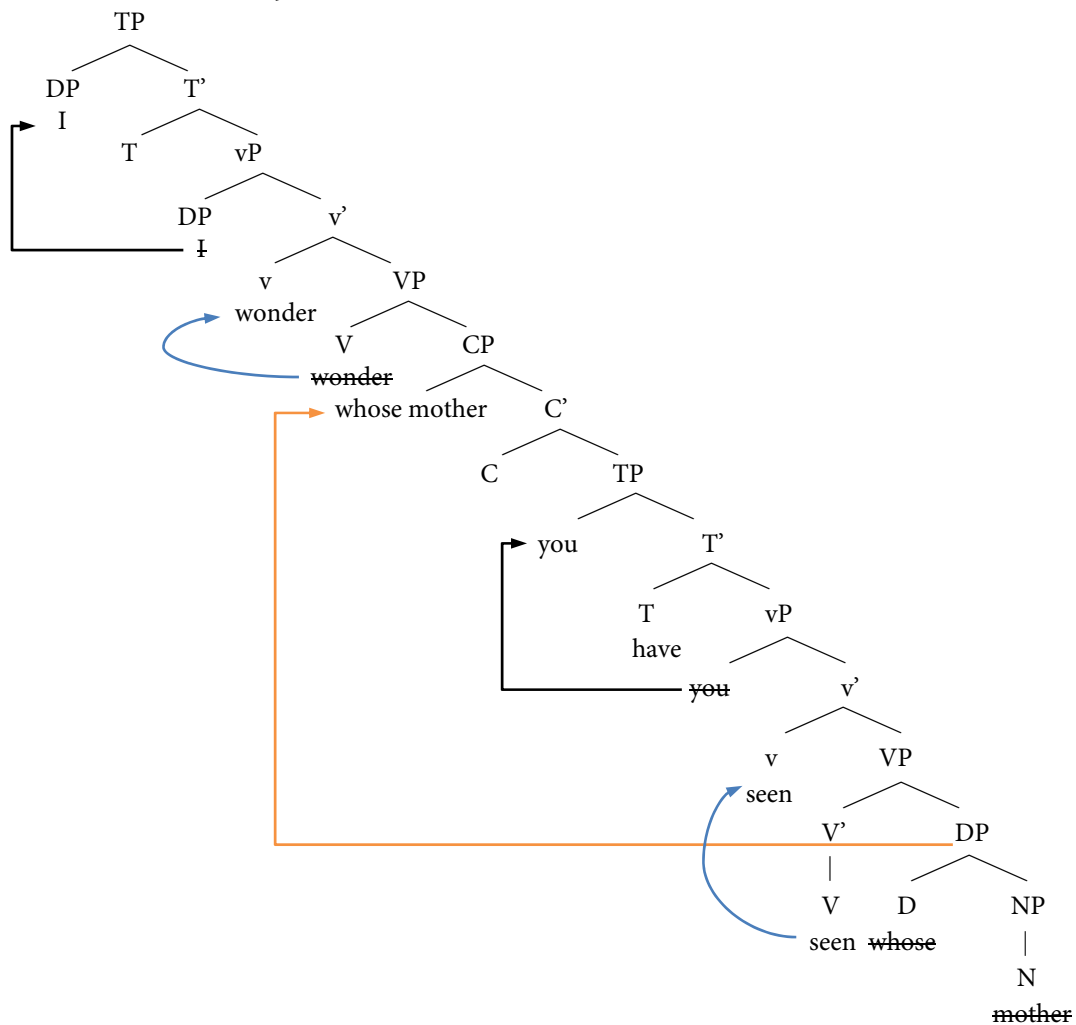
(6)   Given a phrase XP,
        a.   the head X and the specifier YP are pied-pipers for XP;
        b.   complements of X and modifiers (adjuncts) are not pied-pipers for XP.

Webelhuth claims that the *wh*-element has to move to the specifier position of the phrase to be able to percolate its feature to the topmost node.

Webelhuth (1992) claims that PP pied-piping is obligatory in German in questions, but it is optional in English. However, it is problematic in both languages as it does not fit the otherwise observable pattern. Compare (7) and (8).

(7)   a.   *Mit   wem   hat   Hans   gesprochen?*
             With   whom   has   Hans   spoken
             'With whom has Hans spoken?'
        b.   **Wem   hat   Hans   gesprochen   mit?*
              Whom   has   Hans   spoken   with

(8)   a.   *To whom did John talk?*
        b.   *Who(m) did John talk to?*

Webelhuth (1992) brings examples from all Germanic languages showing that it is possible for a PP to be pied-piped in all of them. He does not discuss the optionality of pied-piping in all the languages. This fact leads him to modify his generalization in a way that would yield the required result. He claims that the ability to pied-pipe a phrase is connected to the antecedent being in a theta-marked position or not.

(9)   The antecedent of a constituent in a theta marked position is not a pied-piper.

Cable (2010) challenges the existence of pied-piping as a syntactic operation. One of the most crucial distinctions is between pied-piping as a syntactic operation (10) and pied-piping structures (11).

(10)   Pied-Piping:
        Pied-piping occurs when the operation that targets the feature of a lexical item L applies to a phrase properly containing the maximal projection of L (L^{max}).

(11)   Pied-Piping: Structure:
        A pied-piping structure is one where a phrase properly containing a maximal projection of a *wh*-word (or related operator) has undergone movement typically associated with that operator.

Cable (2010) investigates interrogative sentences and pied-piping in questions. He assumes a Q operator on the lexical items that move and a QP projection in the CP domain. The QP is projected by a phonologically zero Q particle that attracts the lexical item bearing the Q feature. Q is the same operator and feature as *wh* was earlier.
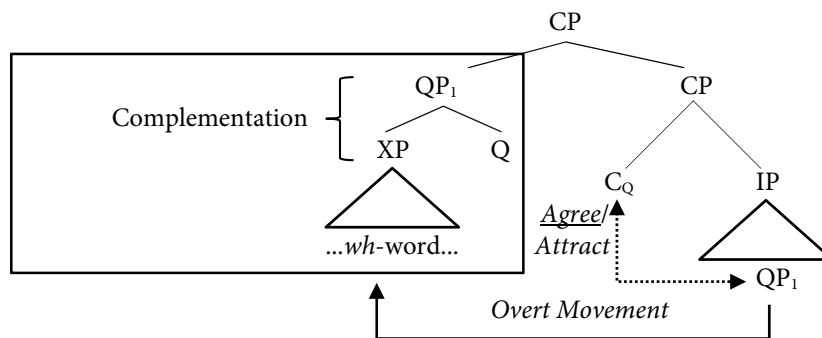
Figure 1: *QP and Pied-piping (Cable 2010:567)*

Cable (2010) argues that there are two types of languages depending on agreement: limited pied-piping languages (12) and non-limited pied piping languages (13).

(12)   Limited Pied-Piping Languages:
       A language where a *wh*-word cannot be dominated in a phrase pied-piped by either an island or a lexical category.

(13)   Non-limited Pied-Piping Languages
       A language where a *wh*-word can be dominated by an island or a lexical category.

Limited pied-piping languages are the ones where there is an Agree relationship between the Q particle and the lexical item bearing the Q-feature. Agreement must be in a sense local in Cable's analysis as well. There cannot be anything between the Q particle and the Q-feature bearing element in the spec of QP (Figure 2).
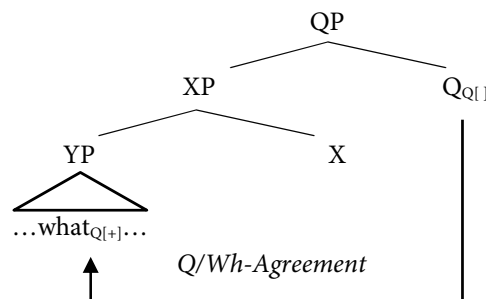
Figure 2: *Q Agreement in English (Cable 2010:583)*

As it can be seen in Figure 2, local agreement is established between the Q-feature and the Q-particle.

(14)   The QP Intervention Principle
       A QP cannot intervene between a functional head F and a phrase selected by F.

Agreement can be blocked if the *wh*-word is embedded in a lexical category deeper in the structure. Cable (2010) assumes a Strong Phase Impenetrability Condition, which means that there can be no agreement relationship between the Q particle and heads in separate Spell-Out domains (Figure 3). Thus constructions like (15) are ill-formed.

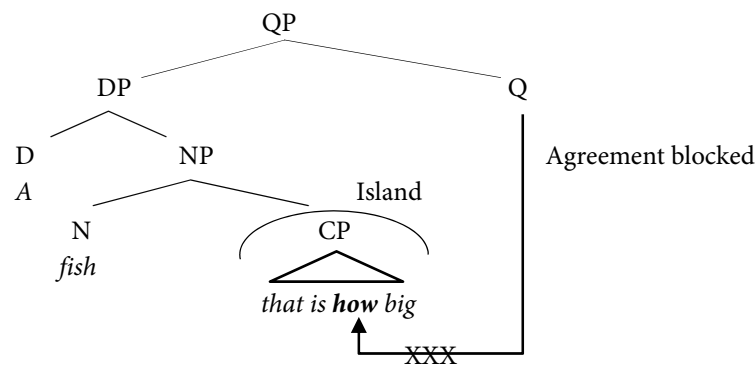(15)   *[*A fish that is **how** big*] *did you buy*?

Figure 3: *Strong Phase Impenetrability Condition (Cable 2010:585)*

3. Focus-movement in Hungarian

As we already saw above, there are languages that have a designated position in a sentence for given information structural functions. Focus is connected to an operator that takes scope over some constituents – it can be narrow, one XP in its domain; or it can be wide, taking a whole predicate in its domain (see in (16)).

(16)   a.   [$_{TopP}$ *Pétert*   [$_{Predicate}$ [$_{Focus}$ *JÁNOS*] *mutatta*   *be*   *Marinak*]].
            Peter-ACC   John   introduced VM Mary-to
            'As for Peter, it was John who introduced him to Mary.'
       b.   [$_{TopP}$ *János* [$_{Predicate}$ [$_{Focus}$ *PÉTERT*] *mutatta*   *be*   *Marinak*]].
            John   Peter-ACC introduced VM Mary-to
            'As for John, it was Peter that he introduced to Mary.'
       c.   [$_{TopP}$ *Pétert*   [$_{Predicate}$ [$_{Focus}$ *MARINAK*] *mutatta*   *be*   *János*]].
            Peter-ACC   Mary-to   introduced VM John
            'As for Peter, it was to Mary that John introduced him.'

The focus is a prepositional operator that can be preceded by Topic Phrase(s). The focus position in Hungarian was first defined as identificational (Kenesei, 1986). The phrase that is moved to this position picks out one referent from a set and identifies it. Horváth (1981) observes the difference between focused and topicalized phrases and formulates a FOCUS-Parameter for every language giving two options for [+focus]-feature (as in (17)).

(17)   FOCUS-Parameter:
        a.   [+FOCUS]: a feature associated freely with any category – deriving the English type languages, that is, Focus in-situ
        b.   the grammaticalized version of the [+FOCUS] feature: an intrinsic part of the feature-matrix of a category, namely V – meant to derive the Hungarian-type, structurally limited, instantiations of focus

This FOCUS-Parameter combines with a Locality Condition on Feature-Assignment – stating that any feature-assigning category must be adjacent to the phrase receiving the feature – can account for the two surface realizations of Focus described in (17). The focus of a sentence is a semantic function identifying a set of items that are exhaustively identified (as in (18)).

(18) The function of focus
The focus represents a proper subset of the set of contextually or situationally given referents for which the predicate phrase can potentially hold; it is identified as the exhaustive subset of this set for which the predicate phrase holds.

This exhaustive identification holds only for structural focus, that is, the immediately pre-verbal position in the structure. Szabolcsi (1981) proposes a test for exhaustivity: whatever one moves to the structural focus position is not exhaustively identified if the subset of the predicate phrase is a logical consequence.

(19) a. *János PÉTERT ÉS ZOLTÁNT mutatta be Marinak.*
        John Peter-ACC and Zoltan-ACC introduced VM Mary-to
        'As for John, it was Peter and Zoltan that he introduced to Mary.'
    b. *János PÉTERT mutatta be Marinak.*
        John Peter-ACC introduced VM Mary-to
        'As for John, it was Peter that he introduced to Mary.'

As it can be seen in (19), (19b) is a logical consequence of (19a), thus (19b) is not an exhaustive identified set, it cannot be exhaustive focus. If we want to interpret the focus exhaustively, then we cannot follow (19a) with a statement of (b), because (19b) does not state exhaustively, who John introduced to Mary. However, if the follow-up sentence (20b) contradicts the predicate in the previous statement (20a), then the focus was exhaustive in it.

(20) a. *János PÉTERT mutatta be Marinak.*
        John Peter-ACC introduced VM Mary-to
        'As forJohn, it was Peter that he introduced to Mary.'
    b. *Nem, Zoltánt is bemutatta neki.*
        no Zoltan-ACC also introduced to.her
        'No, he also introduced Zoltan to her.'

<div align="right">(É. Kiss 2002:79)</div>

Exhaustive identification, however, does not always contrast with a closed set of alternatives. There are examples where the focus alternates with an open set of items, and hence contrast is not there (as in (21)).

(21) a. *A magyar rapszódiákat LISZT FERENC írta.*
        the Hungarian rhapsodies-ACC Liszt Ferenc wrote
        'As for the Hungarian rhapsodies, Ferenc Liszt wrote them.'
    b. *Liszt Ferenc 1886-BAN halt meg.*
        Liszt Ferenc 1886-in died VM
        'As for Ferenc Liszt, it was in 1886 that he died.'

<div align="right">(É. Kiss 2002:80)</div>

The Hungarian pre-verbal focus exclusively exhaustively identifies the set of items it refers to. Szabolcsi (1981, 1983) argues that non-individual-denoting predicates can move to the structural focus position (as in (22)).

(22) a. *János OROSZ LÁNYT vett feleségül.*
        John Russian girl-ACC took as.wife
        'As for John, it was a Russian girl that he married.'

    b. *Péter OKOS LÁNYT akart feleségül venni, nem SZÉPET.*
       Peter smart girl-ACC wanted as.wife to.take not beautiful-ACC
       'As for Peter, it was a smart girl that he wanted to marry, not a beautiful one.'
    c. *János FOKOZATOSAN értette meg a problémát.*
       John gradually understood VM the problem-ACC
       'As for John, it was gradually that he understood the problem.'

(É. Kiss 2002:80)

To sum up, there are reasons to believe that there is a syntactic focus feature in Hungarian that correlates to an exhaustive reading in semantics, and there is a designated, unique syntactic projection corresponding to this discourse function. In what follows, I will present an alternative theory of focus-movement that claims that exhaustive identification is in fact done by an operator in Hungarian (Horváth 1997, 2000, 2005, 2007, 2010).

3.1. Focus-movement is operator movement

Another theory of focus-movement that does not involve a syntactic focus-feature is Horváth (1997 et seq). Horváth (1997) claims that focus-movement is not triggered by a syntactic feature on the focused element, rather there is an operator responsible for the semantics related to focus, that is, responsible for the exhaustive identification reading attached to structural focus in Hungarian. She assumes that there is an operator which she calls EI-Op (Exhaustive Identification Operator) that is attached to a phrase that is associated with the focus interpretation and moves to the CP domain by operator movement. There might be a focus feature on the lexical word that bears main stress – as it is possible in the case of a bigger XP to stress any element inside it.

    Horváth (1997 et seq) claims that the operator, EI-Op bearing a feature [EI] is attached to the focused phrase, and it moves up to the CP domain, where an Exhaustive-Identification Phrase is projected. The movement is triggered by feature-checking, but instead of the lexical element checking its focus-feature, it is the (phonetically vacuous) operator that needs to check its EI-feature (as in (23)).

(23)   The structure for EI-Op movement: (the asterisk indicates the position of main stress)
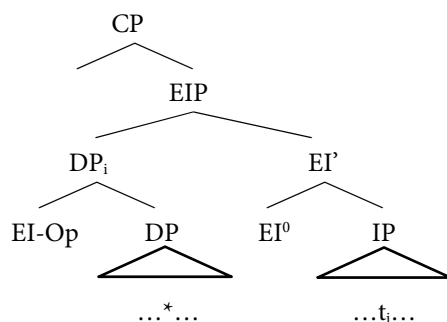
(Horváth 2010:1361)

Horváth (2005) suggests that the prosodic focus can be any constituent contained in the phrase that the EI-Op attaches to (as in (24)).

(24) a. [EI-Op [*MARI*   *Pesten lakó fiát*]]   *hívták*   *fel t.*
               Mary-NOM Pest-on living son-hers-ACC called-3PL up
               'They called up [MARY'S son living in Pest].'

      b. [EI-Op [*Mari*   *PESTEN LAKÓ fiát*]]   *hívták*   *fel t.*
               Mary-NOM Pest-on living son-hers-ACC called-3PL up
               'They called up [Mary's son LIVING IN PEST].'

      c. [EI-Op [*Mari*   *Pesten lakó FIÁT*]]   *hívták*   *fel t.*
               Mary-NOM Pest-on living son-hers-ACC called-3PL up
               'They called up [Mary's SON living in Pest].'

(Horváth 2005:21)

Horváth (1997 et seq) presents a contrast with respect to the restrictions on pied-piping corresponding to movement types. She brings the examples as evidence against a syntactic focus feature. She claims that strong syntactic features cannot pied-pipe a phrase, when the feature-bearing element is embedded inside a pre-nominal adjunct while it is unrestricted with focus-movement, or rather EI-Op movement (as in (25)).

(25) a. * *az ital, amit*   *követelő   vendégektől fél*   *a pincer t*
          the drink which-ACC demanding guests   fear-3SG the waiter
          'the drink customers demanding which the waiter is afraid of…'

      b. * *Mit*   *követelő   vendégektől fél*   *a pincer?*
          what-ACC demanding guests   fear-3SG the waiter
          'Customers demanding what is the waiter afraid of?'

      c. *BARACKPÁLINKÁT követelő*   *vendégektől fél*   *a pincer.*
          apricot-brandy-ACC   demanding guests   fear-3SG the waiter
          'It is customers demanding APPRICOT BRANDY that the waiter is afraid of.'

Horváth claims that the insensitivity of focus to pied-piping is due to the fact that the operator is situated outside the phrase, and thus, Agree between the [EI]-feature and the EIP in CP is not blocked by the ph(r)ase boundary.

     To sum up, the structural focus position is associated with an exhaustive semantic reading that can be accounted for in various ways according to the above-mentioned theories. The existence of the syntactic focus-feature has been questioned by several authors (Horváth 1997, 2000, 2005, Zubizarreta1998 among others). Those theories suggest a discourse related feature or operator to account for the exhaustive reading of focus. One of the main goals of this paper is to experimentally test the behavior of the focus-construction in pied-piping, which can hopefully shed some light on the nature of the focus-feature and provide further evidence for one approach or the other.

3.2. Wh-*movement in Hungarian*

Hungarian is a *wh*-fronting language, which means that the interrogative pronoun or phrase must move to a position that is structurally high to check its [wh]-feature. *Wh*-phrases target a position in the CP domain, the same position of focused elements – that is, *wh*-phrases move to FP, as they bear the feature [+focus] additionally to the [wh]-feature (among others Horváth 1986, É. Kiss 2002). É. Kiss (2002) considers *wh*-phrases to be inherently focused, hence they

move to FP[1]. The reason for *wh*-phrases having a [+focus]-feature as well is that they require an exhaustive answer (as in (26)). The role of exhaustive identification is associated with focus in Hungarian.

(26) a. **János* [_AspP_ *be mutatott kit Marinak*]?
      John   VM introduced whom Mary-to
      'Whom did John introduce to Mary?'
    b. *János* [_FP_ KIT$_i$ [_VP_ *mutattot be* t$_i$ *Marinak*]]?
      John   whom   introduced VM   Mary-to
      'Whom did John introduce to Mary?'

If there is a focus and a *wh*-phrase in the same sentence then only one of them can move to the specifier of FP, and that has to be the *wh*-phrase (see (27)).

(27) a. **CSAK PÉTERT látta KI?*
      only   Peter-ACC saw who
      'Who saw only Peter?'
    b. *KI látta CSAK PÉTERT?*
      who saw only   Peter-ACC
      'Who saw only Peter?'

The *wh*-phrase has to move to spec, FP presumably for semantic reasons – the *wh*-phrase is only interpreted as a question if it combines with a [+focus]-feature and moves to the specifier of FP (to check its features). The focus in the case of (27) has been marked by the phrase *csak* 'only', which is a focus particle in Hungarian. It can elicit the focus reading without having to occupy the specifier of the focus projection, FP. There can even be two *csak*-phrases and a *wh*-phrase in a sentence (as in (28)).

(28) a. *MELYIK FÉLÉVBEN* [_VP_ *kapott CSAK HÁROM LÁNY CSAK KÉT TÁRGYBÓL JELEST*]?
      which   term-in   received only three   girl   only two subject-from A+
      Reading 1: 'In which term was it only three girls who received an A+ only in twosubjects?'
      Reading 2: 'In which term was it only in two subjects that only three girls received an A+?'
    b. *MELYIK FÉÉLÉVBEN* [_VP_ *kapott CSAK KÉT TÁRGYBÓL JELEST CSAK HÁROM LÁNY*]?
      which   term-in   received only two subject-from A+   only three   girl
      Reading 1: 'In which term was it only in two subjects that only three girls received an A+?'
      Reading 2: 'In which term was it only three girls who received an A+ only in two subjects?'

In the case of a sentence as in (28), in which there are two *only*-phrases and a *wh*-phrase, the *wh*-phrase moves to the specifier position of FP overtly, and the two only phrases stay inside the VP. The fact that both readings are available with both word orders proves that the *only*-phrases stayed in-situ inside the flat VP and they mutually c-command each other, hence their relative scope with respect to each other is equal. That is, the scope of the *only*-phrases can inform us about the structural positions they take inside the clause, if there was a fixed order one taking scope over the other, it would suggest that one is in a higher – maybe adjoined – position inside the clause. The inner structure of the verb phrase is flat/non-hierarchical in

---

[1] The Focus Phrase first was labeled FP (Brody 1990a), referring to the fact that it is a functional projection and later was more specifically labeled FocP (É. Kiss 1998).

Hungarian, and thus the constituents inside the VP mutually c-command each other. Scope taking is governed by c-command – the constituent XP that c-commands another constituent YP takes scope over it. YP can take scope over XP if it moves covertly to a higher position in LF, which is not the case in (28).

As stated above, Horváth (1986) assumes that every *wh*-word that moves up to C has to bear a [+focus] feature. She makes this claim based on the fact that the movement of the *wh*-phrase triggers the inversion of the verb modifier and the verb (as in (29))– just like in focus.

(29) a. [_{TopP} *A   huzat* [_{FP} *MELYIK SZOBA ABLAKAIT   törte   be*]]?
        the draft   which   room's windows-ACC broke in
        'The windows of which room did the draft break?'
    b. [_{FP} *MELYIK SZOBA ABLAKAIT   torte   be   a   huzat*]?
        which   room's windows-ACC broke in   the draft
        'The windows of which room did the draft break?'
    c. * *Melyik szoba ablakait   a   huzat* [_{AspP} *be törte*]?
        which   room's windows-ACC the draft   in broke

Embedded questions in Hungarian contain the same interrogative pronoun, but they are introduced with the complementizer *hogy* 'that' (30). É. Kiss (2002) argues that in embedded questions a need arises for a separate ForceP projection as the [+/- wh]-feature has to be encoded somewhere in the structure, and the complementizer *hogy* 'that' usually does not have a [wh]-feature.

(30) *János meg kérdezte,* [_{CP} *hogy* [_{TopP} *Pétert*   [_{FP} *ki   mutatta   be   Marinak*]]].
    John VM asked   that   Peter-ACC   who introduced VM Mary-to
    'John asked who introduced Peter to Mary.'

                                                                       (É. Kiss 2002:99)

To sum up, *wh*-phrases in Hungarian have to be moved out of VP to a structurally higher position in the CP domain (e.g. CP, FocP, ForceP). The movement of the *wh*-phrase is either triggered by a focus-feature on the *wh*-phrase or by some other morpho-syntactic requirement. This morpho-syntactic feature can be [wh]-feature or as Cable (2010) suggests there is a Q-operator that attaches to the phrase that bears a [wh] feature. Q has its own feature that it needs to check in the CP domain, that is why it moves to CP.

4. Discourse Linking

Discourse-linking has been considered in connection with *wh*-movement since Pesetsky (1987). It has been observed that there is a difference among wh-phrases with respect to the availability of movement out of syntactic islands. Pesetsky (1987) claims that discourse-linked *wh*-phrases ask a question about something that is part of a set of referents, which are pre-established in the discourse or are part of shared knowledge. That is, the entity the *wh*-word poses a question about has a set referent in the discourse.

(31) [Which boy] did you invite to the party?

Discourse-linking is important in cases when the *wh*-element should move out of a syntactic island. Syntactic islands can be grouped into different categories based on their ability to allow movement out of the island. There are weak islands and strong islands: weak islands allow *wh*-phrases to move out of their boundary, while strong islands prohibit any type of movement out of them.

As it can be seen in (32), movement out of an island is acceptable when the *wh*-word is discourse-linked, whereas the non-discourse-linked *wh*-word '*how*' cannot be moved out of the embedded question.

(32) a. Which book did John wonder [whether to read ____]?
      b. *How did John wonder [whether to read a book ____]?

Discourse-linking is an important part of the experiments presented in this paper. I take discourse-linking as a feature of phrases. A discourse-linked phrase has a referent in a set that is known or part of some common knowledge of the speakers. In this way, I take *wh*-phrases to be either discourse-linked – the ones that have set referents in the discourse, or non-discourse-linked – the ones such as '*how*' and '*why*', which are open as to the possible referents in discourse. It is unconventional to call phrases that are non-interrogative non-discourse-linked, however, I believe that the phrases answering a non-discourse-linked *wh*-phrase can be considered non-discourse-linked. In the experiments presented in the next section, I put focused phrases into the discourse-linked and non-discourse-linked group based on what type of *wh*-phrase they answer.

5. Experiments

The following experiments were created to answer the following research questions:

Q1: Is pied-piping permissible in *wh*-constructions and focus-constructions?
Q2: Is discourse-linking a factor in the availability of pied-piping?

A starting point of these experiments was a set of examples given by Horváth (1997) as a proof for the non-syntactic nature of the focus-feature by contrasting the pied-piping behavior presented in *wh*-movement triggered by a syntactic feature and focus-movement. However, a pilot study conducted earlier suggested that pied-piping might be acceptable in *wh*-constructions as well. I conducted two experiments, one that contained pied-piping by *wh*-expression embedded in a prenominal adjunct (Experiment 1) and one where pied-piping was triggered by the focused element embedded in a prenominal adjunct (Experiment 2).

5.1. Experiment 1 – WH

This experiment was based on the pilot study. Based on feedback on the experiment, I decided to change the target sentences. The number of factors was reduced to make a 2×2 design. This made for less statistical comparisons and a stronger statistical model.

5.1.1. Method

5.1.1.1. Subjects

30 adult Hungarian native speakers participated in the experiment. The majority of the participants were students of the University of Debrecen.

5.1.1.2. Procedure

The experiment was built and run in Ibex Farm (www.spelout.net). The subjects were presented with the target sentences one by one on the screen. The link to the experiment was sent out via email, and every subject did the experiment online. This experiment was an Acceptability Judgment Task, the subjects had to judge each sentence on a 7-point Likert-scale.

5.1.2. Materials

In this experiment there were only sentences containing pied-piping in *wh*-construction. We tested pied-piping by a *wh*-phrase embedded in a prenominal adjunct. We investigated pied-piping by non-discourse-linked *wh*-phrases (as in (33)), and pied-piping by discourse-linked *wh*-phrases (as in (34)). The *wh*-phrase was embedded inside a DP that had a definite determiner on the left edge of the phrase. The baseline sentences were neutral sentence containing no movement inside the embedded clause. The target sentences involved *wh*-pied-piping in the embedded questions. The questions are embedded under predicates that require an embedded interrogative clause, such as: *megkérdeztem* 'I asked', *nem tudom* 'I don't know', *fogalmam sincs* 'I have no clue' and *érdeklődtem* 'I inquired about something'. There were 32 target sentences: 8 without pied-piping with an expression that can be considered to have no specific referent/set of referents, 8 with pied-piping by a non-discourse-linked *wh*-phrase; 8 sentences without pied-piping containing an expression that can be considered to be discourse-linked in the sense that it has a specific referent associated with it, 8 corresponding sentences with pied-piping by discourse-linked *wh*-phrases. In each condition, the sentences without pied-piping served as baseline sentences to the ones with pied-piping. The 32 target sentences were divided into two lists – one list contained either the baseline or the pied-piping version of the pairs.

(33) Baseline (DP in post-verbal position):
  a. *Azt   hallottam, hogy az HBO filmet   forgatott a*
      that herad-1SG that the HBO film.ACC shot-3SG the
      tömeggyilkosságért letartóztatott bűnözökről tavaly.
      mass.murder.for incarcerated criminals   last.year
      'I heard that the HBO was shooting a movie about the criminals incarcerated for mass murder last year.'

  Pied-piping: … [DP D [NP [WH $_{obl}$ participle] N $_{acc}$] V VM ADV
  b. *Nem tudom,   hogy a   miért letartóztatott bűnözőkről forgatott filmet   az HBO tavaly.*
      not know-1SG that the why incarcerated criminals shot-3SG film.ACC the HBO last.year
      'I don't know the HBO shot a movie about the why incarcerated people.'

(34)   Baseline (object in post-verbal position):

a.   *Úgy   tudom,   hogy az   építészkamara   kizárta   a*
       so   know-1SG that   the   architect.union banned-3SG the
       kartonpapírból   készített modelleket   a   tervpályázatból.
       cardboard.out.of made   models.ACC   the   plan.tender.from
       'I believe the architects union has banned the models made of cardboard from the tender.'

Pied-piping: … [DP D [NP [WH obl participle] N acc] V VM ADV

b.   *Fogalmam sincs, hogy a   miből   készített modelleket*
       my.clue   not   that   the what.out.of made   models.ACC
       zárta   ki   az   építészkamara   a   tervpályázatból.
       banned   VM   the   architects.union the   plan.tender.from
       'I have no clue as to made out of what models the architects union has banned from the tender.'

5.1.3. Results

Preliminary statistical analysis of the data shows a tendency for degradation rather than a clear difference between the baseline and target sentences. The results were analyzed with linear mixed models – however, the model fitting needs further tests to find the best model that fits the data. Figure 4 shows the results of the experiment.
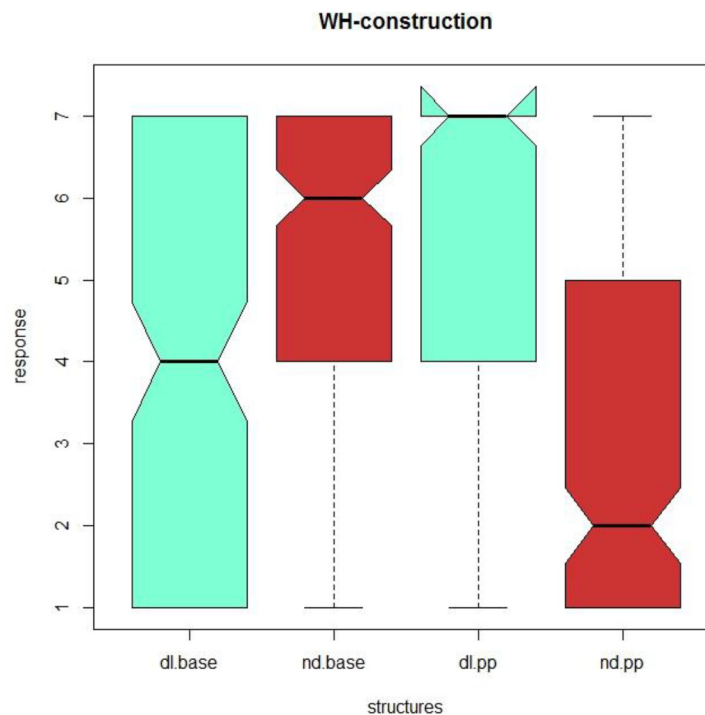

Figure 4: *Pied piping in wh-movement*

As it can be seen in Figure 4, pied-piping is acceptable when the pied-piper is a discourse-linked *wh*-phrase, whereas pied-piping is unacceptable when the pied-piper is a non-discourse-linked *wh*-phrase.

5.2. Experiment 2 – FOC

In this experiment we tested pied-piping by focus structures. This experiment, just like the one above, is a separate test to make sure that the subjects do not award low points for the structures because their working memory is full and they cannot pay attention to the task.

5.2.1. Method

5.2.1.1. Subjects

The experiment was done by 32 adult native Hungarian speakers, most of whom attend the University of Debrecen.

5.2.1.2. Procedure

The experiment was built and run in Ibex Farm (www.spellout.net). The experiment was an Acceptability Judgment Task, in which the subjects judged the sentences on a 7-point Likert-scale.

5.2.2. Materials

There were 48 test sentences altogether: 32 target sentences and 16 filler sentences. The number of filler sentences is only half of the target sentences because the target sentences were divided into two lists – hence each list contained 32 test sentences – 16 target and 16 filler sentences. The target sentences contained 16 discourse-linked expressions (as in (35)) and 16 non-discourse-linked expressions (as in (36)). It might seem unusual to classify expressions that are not *wh*-phrases into discourse-linked and non-discourse-linked groups. However, the way I understand discourse-linking it means that there is a (set of) specific referent(s) that the expression is associated with. This way, phrases that are not associated with such sets are considered non-discourse-linked. From the 32 (16/16) target sentences half of them did not contain movement, that is, the object DP stayed in-situ in its post-verbal position. In the target sentences with pied-piping, the pied-piper is embedded in a focused constituent and it undergoes movement to the pre-verbal focus-position.

(35) Baseline (DP in post-verbal position):
 a. *Azt hallotta, hogy a biztosító megvédi a*
  that heard-1SG that the insurance.company VM.protect-3SG the
  téglából épített épületeket tetőbeázás esetén.
  brick.from built buildings.ACC roof.leaking case.in
  'I heard that the insurance company protects buildings made of brick in the case of a roof leak.'

 Pied-piping: …[[FOC_obl participle] N_ACC] NP V VM ADV
 b. *Furcsálltam, hogy csak a téglából épített épületeket védi*
  weird-1SG that only the brick.from built buildings.ACC protect-3SG

meg   a   biztosító   tetőbeázás   esetén.
VM   the   insurance.company   roof.leaking   case.in
'I found it weird, that it was only the building made of brick that the insurance company protects in the case of a roof leak.'

(36)   Baseline (DP in post-verbal position):
   a.   *Azt   hallottam,   hogy   az   ételkritikus   megdicsérte   a   magyarosan*
        that   heard-1SG   that   the   food.critic   VM.praised-3SG   the   Hungarian-style
        fűszerezett   ételeket   a   múlt   heti   cikkében.
        spiced   dishes.ACC   the   last   week   article.his.in
        'I heard that the food critic praised the dishes made with Hungarian-style spices in his article last week.'

   Pied-piping: …[[FOC_obl participle] N_ACC] NP V VM ADV
   b.   *Csodálkoztam,   hogy   csak   a   magyarosan   fűszerezett   ételeket*
        surprised-1SG   that   only   the   Hungarian-style   spiced   dishes.ACC
        dicsérte   meg   az   ételkritikus   a   múlt   heti   cikkében.
        praised-3SG   VM   the   food.critic   the   last   week   article.his.in
        'I was surprised that it was only the dishes made with Hungarian-style spices that the food critic praised in his article last week.'

5.2.3. Results

Three subjects had to be excluded from statistical analysis since they did not use the scale appropriately – they used only one value of the scale to all test items. The results show that pied-piping is acceptable in focus movement. Although there are specific items that some subjects rated lower than expected, the degradation is not refined enough to show in statistical analysis. The results were analyzed by linear mixed models, and show no significant difference between the baseline and the target condition (see Figure 5).
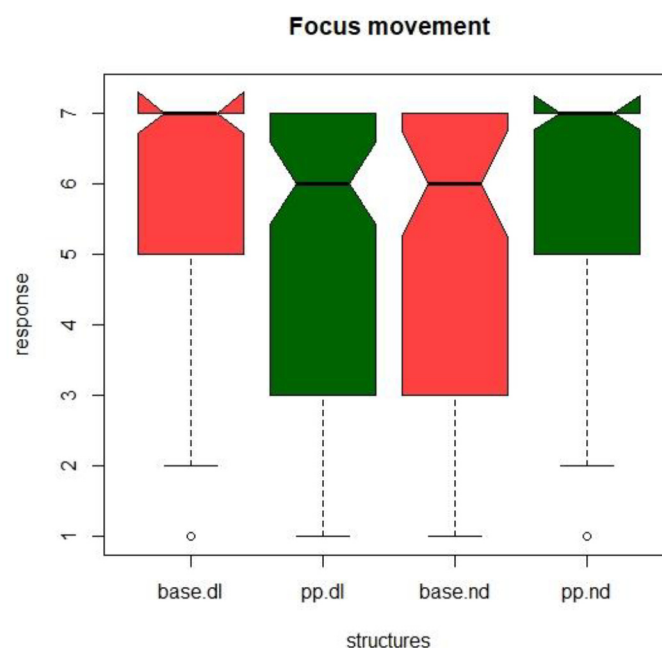

Focus movement

Figure 5: Pied-piping in focus-movement

The diagram shows that pied-piping is acceptable in the discourse-linked and non-discourse-linked conditions as well. As mentioned above (in section 4), discourse-linking means that the focused phrase is an answer to either a discourse-linked or a non-discourse-linked *wh*-phrase.

6. Discussion

6.1. *Focus-feature*

Pied-piping in Hungarian focus-structures is assumed to be unrestricted (Horváth 2000, 2005). The experiments conducted throughout this research seem to support the idea of the unrestricted nature of pied-piping in focus-constructions. The pilot study followed an experimental practice in which the baseline sentences themselves already contained focus-movement – that is, the baseline sentence already was of a type of structure that is different from a neutral sentence. This design was changed in the experiments presented here: baseline sentences were constructed which did not have any type of movement in them. The results of that did not give statistically significant degradation in the non-discourse-linked condition, however, the ratings were lower than what we found in the pilot studies. These results might suggest that the focus-feature is not encoded in syntax, only in prosody and semantics/pragmatics. As the presence of a syntactic feature makes pied-piping less acceptable, we assume that the acceptability of pied-piping in focus constructions is a result of the lack of a syntactic focus-feature.

6.2. Wh-*feature*

The experiments have confirmed that pied-piping by a pre-nominal adjunct in *wh*-movement is as acceptable as it is in focus-movement. (cf. Horváth 2000) These results suggest that the *wh*-feature is similar in its nature to the focus-feature: neither feature acts as a trigger for syntactic movement in Hungarian. This finding could be accounted for by theories of *wh*-movement in languages where the *wh*-elements move to a designated position in the sentence that assume that there is a separate operator (a Q operator) that is responsible for the semantics and the syntactic reordering of interrogative sentences, such as Cable (2010). The picture, however, is not that clear in Hungarian.

6.3. *Proposal*

The results of the experiments paint an interesting picture regarding the pied-piping behavior of Hungarian. Though pied-piping seems to be unrestricted in focus-movement, it seems to be split in two regarding *wh*-movement. Pied-piping by a discourse-linked *wh*-element is unrestricted similarly to focus- (2010) we can argue that there are two different types of *wh*-elements in Hungarian: (i) discourse-linked *wh*-phrases that do not need to *Agree* (Figure 6), and (ii) non-discourse-linked *wh*-phrases that do need to *Agree* (Figure7).
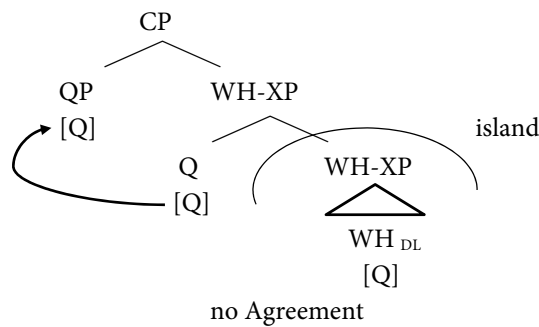
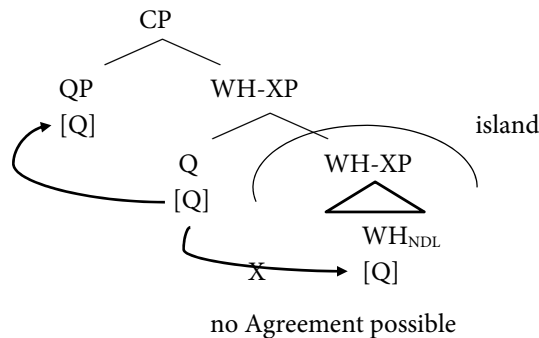*Figure 6: No Agreement between wh-word and Q head*

Figure 7: Agreement needed but blocked

7. Conclusion

This paper presented two experiments investigating pied-piping by prenominal adjuncts in Hungarian. Section 2 gives an overview of two relevant theories on pied-piping. Section 3 presents the literature background on Hungarian focus-movement and *wh*-movement. Section 4 introduces discourse-linking. Section 5 presents the experiments and the results of the two experiments. It also contains a tentative proposal for the pattern drawn by the results of the experiments. Based on the results we can claim that the focus-feature in Hungarian is inactive with respect to syntax. It might be active in prosody and it is active in semantics. The case of the *wh*-feature is more complicated. The nature of the *wh*-element makes a difference in its ability to allow pied-piping. Syntactic Agreement is necessary for non-discourse-linked *wh*-phrases between the Q-head and the *wh*-element, while discourse-linked elements do not need to establish an Agree relationship with the Q head.

References

Cable, S. 2010. *The Grammar of Q: Q-Particles, Wh-movement and Pied-piping.* Oxford: OUP.

Horvath, J. 1981. Aspects of Hungarian Syntax and the Theory of Grammar. PhD Diss. UCLA.

Horvath, J. 2000. Interfaces vs. the Computational System in the Syntax of Focus. In H. Bennis, M. Everaert and E. Reuland (eds), *Interface Strategies*, 183–206. Amsterdam: Royal Netherlands Academy of Arts & Sciences.

Horvath, J. 2005. Separating 'Focus Movement' From Focus. Manuscript. Tel-Aviv University.

Kenesei, I. 1986. On the logic of Hungarian word order. In W. Abraham and S. de Meij (eds.), *Topic, Focus and Configurationality*, 143–159. John Benjamins Publishing.

É. Kiss, K. 1998. Identificational Focus versus Information Focus. *Language* 74(2): 245–273.

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

Lasnik, H., and Saito M. 1992. *Move α. Conditions on Its Application and Output*. MIT Press, Cambridge, Mass.

Pesetsky, D. 1987. Wh-in-situ: Movement and unselective binding. In E. J. Reuland, A. G. B. Ter Meulen (eds.), *The representation of (in) definiteness*, 98–129. Cambridge, MA: MIT Press.

Szabolcsi, A. 1981. Compositionality in focus. *Folia Linguistica* 15: 141–162.

Szabolcsi, A. 1983. The Possessor that ran away from home. *Linguistic Review* 3: 89–102.

Webelhuth, G. 1992. *Principles and Parameters of Syntactic Saturation*. Oxford/New York: Oxford University Press.

The British are coming…! A bibliometric analysis of L2 vocabulary research in 1988

Paul Meara

Swansea University, United Kingdom

Abstract

This paper uses a co-citation analysis to examine the research on L2 vocabulary acquisition that was published in 1988. Two analyses are presented. The first is a detailed account of the 1988 research on its own terms. The second analysis places this work in a larger context by looking at the research published in a five-year window covering 1984–88. The analyses identify important themes in the research and significant sources who are influencing the way the research is developing at this time. A particularly important new research theme centred around corpus linguistics appears in the 1988 data, and there are some surprising changes to the list of influential sources.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric analysis

1. Introduction

This paper is the eighth in a series of studies that 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 in 1982, 1983, 1984, 1985, 1986, 1987 and 2006 (Meara 2012, 2014, 2015, 2016, 2017, 2018 and 2019). This paper is a sequel to my earlier LingBaW papers, in that it covers the research output of 1988. The paper contextualises this research by means of an additional analysis of the research published in the five-year window of 1984–88.

The paper falls into two parts. Part I reviews the new research that appeared in 1988 in its own terms. Part 2 provides a wider context for this research, by summarising the main trends that appear in a five-year window covering 1984–1988. Both parts make use of the co-citation methodology that was used in the earlier papers. For readers who are not yet familiar with this approach, I have provided a methodological summary in an Appendix.

2. Part 1. The new research published in 1988

At first glance, 1988 seems to be a poor year for vocabulary research. The previous four years had shown a small but steady increase in the number of research outputs published, but for 1988 the total number of outputs falls just below 100 for the first time since 1985. However, on

closer inspection, this fall turns out to rather deceptive. The Vocabulary Acquisition Research Group Archive (VARGA) database for 1988 (Meara n.d.) identifies only 99 outputs, but the make-up of these sources is rather different from what we found in earlier years.

2.1. The data sources

In 1988, we have a relatively large number of books (9) and a handful of computer programs (2), as well as a grand total of 21 book chapters that deal with L2 vocabulary acquisition in one way or another, and 50 standard research papers. The VARGA database for 1988 also records five PhD theses, and three cited Masters theses. However, since the research literature tends not to cite theses reliably, VARGA typically under-reports work of this type, and it is possible that some sources of this type have been missed. VARGA includes theses which are cited in later research work, but does not attempt to log every thesis presented, and the work listed here should not be taken as an exhaustive list. Not all of the work published in 1988 is included in the analysis that follows. Books, theses and similar works are conventionally excluded as raw data in bibliometric analyses of the type reported in this paper, on the grounds that they cite the research in a way which is different from what we expect in a normal research paper. A few works of this type have been excluded from the analyses reported in this paper. I have also excluded from the analysis two computer programs, developed by Esser & Widdig and by Scott & Johns. Both these works are reported more fully in separate research papers which *are* included in the analysis. Table 1 lists the complete set of 1988 publications which are excluded from the analysis that follows.

Table 1: *Work published in 1988 but excluded from the analysis in this paper*

BOOKS

Boch, R. 1988. *Les faux amis aux aguets: Dizionario di false analogie e ambigue affinità fra francese e italiano*. Bologna.

Broeder, P., G. Extra, R. van Hout, S. Stromqvist and K. Voionmaa. 1988. *Processes in the developing lexicon*. Tilburg.

Browne, V. 1988. *Odd pairs & false friends: Dizionario di false analogie e ambigue affinità fra inglese e italiano*. Bologna.

Carter, R. and M. McCarthy (Eds.) 1988. *Vocabulary and Language Teaching*. London: Longman.

Leiste, D., C. Döll and A. M. Tereso Domingos. 1988. *Kleines Wörterbuch der 'falschen Freunde': Deutsch-Portugiesisch, Portugiesisch-Deutsch*. Leipzig.

Lozanov, G. and E. Gateva. 1988. *The Foreign Language Teacher's Suggestopedic Manual*. Gordon and Breach Science Publishers.

SOFTWARE

Esser, R. and W. Widdig. 1988. *WE/WT, Autorensystem fur die Wortschatzarbeit. Benutzungsanleitung*. Cologne: RRZK.

Scott, M. and T. Johns. 1988. *Oxford English Software: Microconcord 1.0*. Oxford: Oxford University Press. 1988.

THESES and DISSERTATIONS

Ahmed, M. O. 1988. Vocabulary Learning Strategies. PhD Thesis, UCNW Bangor. 1988.

Mei Lin. 1988. An assessment of the treatment of vocabulary in a series of pilot coursebooks used by English-Language learners at tertiary level in the Peoples' Republic of China. MA Thesis. York University.

Novda, L. M. 1988. The word retrieval process and reading acquisition and development in bilingual children. PhD thesis. Harvard University.

Sonaiya, Q. C. 1988. The lexicon in second language acquisition: a lexical approach to error analysis. PhD thesis. Cornell University.

Söderman, T. 1988. Word associations of foreign language learners and native speakers – a shift in response type and its relevance for a theory of lexical development. Masters Thesis. Åbo Akademi.

Verkaik, P. and P. van der Wijst. 1988. Taal verlies en woordherkenning in het Frans als vreemdetaal. [Language loss and word recognition in French as a foreign language.] Masters thesis: Katholieke Universiteit, Nijmegen.

Weltens, B. 1988. The attrition of French as a foreign language. PhD Thesis, Katholieke Universiteit Nijmegen.

White, W. H. 1988. Vocabulary acquisition from reading. PhD dissertation, University of Southern California, Los Angeles.

The books and monographs listed in Table 1 are particularly interesting. Three of these are studies that deal with cognates and false friends: Boch (1988) and Browne (1988) both treat false friends in French and Italian; Leiste, Döll and Tereso Domingos (1988) treats false friends in German and Portuguese. Studies of this type have appeared occasionally in previous years, but this set is larger than most, and extends the range of languages. Lozanov and Gateva (1988) deal with Suggestopedia – the idea that languages can be learned subconsciously. This book is not specifically concerned with vocabulary, but vocabulary plays a large role in studies of subconscious language learning, and for this reason, it has been included in the VARGA listing.

Two of the book-length studies are substantial project reports. Broeder et al (1988) summarise the on-going work of the European Science Foundation Project as far as the lexicon is concerned (Purdue 1982, 1993). The report consists of 5 main chapters dealing with: the learner's use of words; the informants and data base used in the studies; richness and variety of the developing lexicon; pronominal reference to person; and word formation processes in talking about entities. Coenen and Vermeer (1988) is another substantial study that attempts to assess how much L2 vocabulary is known by migrant learners of Dutch. Part One of this volume describes how the basic word lists were collected. Part Two consists of a list of 4332 different words used by more than one of the speakers tested.

The rationale for excluding book length treatments is that their normal citation practices differ from what we find in standard research papers. However, the final book, Carter and McCarthy (1988), is odd in this respect. This volume is an edited collection of ten papers, four of which were written by the editors. After some thought, I decided to treat this volume as ten separate entries, despite the overwhelming contribution of the editors.

A new development in the 1988 outputs is the appearance of two computer programs. The most important of these is Scott and Johns' *Microconcord*, a concordancing program. This program was unusual at the time in that it ran on small microcomputers, and did not require a mainframe computer. It also had a very short learning curve, which meant that it could readily be used in ordinary classroom situations. It was rapidly taken up as a tool that language teachers could use to help their students learn the finer points of vocabulary. I excluded this item from the analysis that follows, treating it as a book, rather than a paper. *Microconcord* is not completely neglected, however: it is described and evaluated in other papers which are included in the analysis. The second computer program (Essen and Widdig 1988) appears to be an authoring program for developing computer-based vocabulary exercises. This work was not included in the analysis. It is described in another paper by the same authors, and this paper is included in the data set.

Table 2 lists the small number of papers which I was not able to source, mainly due to the closure of libraries and restrictions on travel during the coronavirus pandemic.

Table 2: *Papers published in 1988 which were unobtainable, and were consequently excluded from the analysis*

Armand, E. (1988). Typologie des exercices de vocabulaire contenus dans les manuels de français édités en France entre 1970 et 1984 et destinés a l'enseignement-apprentissage du français aux étrangers. [A typology of vocabulary exercises found in French course-books.] *Bulletin de l'Unité de Recherche Linguistique* 4: 97–183.

Galisson, R. (1988). Le vocabulaire revient: Le vocabulaire en pénitence. Brève histoire d'une disgrace chronique. *Reflet* 27: 14–19.

Noyau, C. (1988). Le développement du lexique en langage étrangère: lemmatisation de données orales d'acquisition. [The development of the lexicon in a foreign language: analysis of spoken language data.] *Actes du 2ème Colloque de Linguistique Hispanique*, 199–216. Brest: Université de Bretagne Occidentale.

Taylor, A. (1988). Learners and English dictionaries: some assumptions and challenges. *Institute of Language in Education Journal* 4: 88–92.

Tréville, M-C. (1988). Faut-il enseigner le vocabulaire de la langue seconde? [Do we need to teach vocabulary in a second language?]. In R. LeBlanc, J. Compain, L. Duquette and H. Séguin (eds.) *L'enseignement des langues secondes aux adultes: recherches et pratiques*. Ottawa: Presse de L'Université d'Ottawa.

The four French papers in this table are probably an important omission. We have seen in our earlier papers that French research on vocabulary was beginning to go through something of a resurgence in the late 1980s, and the omission of these papers means that this strand of research is not properly represented in the main analysis. The papers by Armand and Noyau are technically available, but they could only be found in libraries that were not functioning because of the corona virus pandemic. The paper by Galisson was completely untraceable. Again, I suspect that this is an important omission, as Galisson was a very vocal critic of vocabulary teaching in France in the 1980s, and his work developing an alternative pedagogy of vocabulary and lexis was both influential and very controversial. Tréville is part of a vocabulary research group based in Ottawa which was developing an approach to the teaching of French vocabulary that was more empirical and less philosophical than the approaches being developed in France at that time.

The paper by Taylor proved to be untraceable: the paper seems to be a large scale study of dictionary use by L1 Chinese learners of English at University level, and would have made a useful addition to the other research on dictionary use which appeared in 1988.

The remaining 81 items, all conventional book chapters or papers in journals, are not listed here in full for reasons of space. Readers who want to access this list can do so by using the VARGA database: https://www.lognostics.co.uk/varga/ and entering the search term **## 1988.**

The usual superficial analysis of this data set identifies 92 unique contributors. As usual, we find that most of these authors contribute to only a single paper in the data set, and only a handful of authors contribute to more than one paper. The data is summarised in Table 3.

Table 3: *Authors contributing to N papers in the main data set*

No of papers (N)	5	4	3	2	1
Actual data	0	4	1	4	83
Lotka's model N=83	3	5	9	20	83

Carter, McCarthy, Meara and Palmberg all contribute 4 entries, making them the most prolific authors in this year's output. McCarthy is a new entrant to the prolific authors list. Johns, also a new entrant, contributes three papers. Four authors, Nation, Robinson, Summers and Tono all contribute two papers to the data set. Of these, only Nation has appeared in previous counts.

Table 3 also shows that this profile is considerably shallower than we would expect to find in a mature research field. With 83 authors contributing just one paper, we would expect to find about 20 authors who contribute to two papers, nine who contribute to three papers, and so on. In 1988, the field is still short of contributors with a substantial output, just as it was in the previous years. (Lotka 1926. For readers who are unfamiliar with Lotka's approach, a brief account can be found in Appendix 1.).

2.2. The analysis

The main analysis in this paper is a co-citation analysis of all the sources cited in the 1988 data set. The methodology is described in Appendix 2. This analysis identified 1391 authors who are cited in the data set, a substantial fall on the 1987 figure of 1587 that I reported in the previous paper. The number of times each of these sources is cited in the data set is summarised in Table 4.

Table 4: *The number of times sources are cited in the 1988 data set.*

frequency	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	1	1	1	1	3	3	2	4	5	10	12	39	55	186	1068

The most cited sources in this data set are Sinclair (15) Nation (14), Meara (13), Channell (12), Krashen, Ostyn and Renouf (11), Carter, McCarthy and Richards (10), and Halliday and West each cited in 9 papers. The main point to note here is the emergence of Sinclair as the most cited source in the data set. Allowing for the smaller size of the 1988 data set, these figures are very comparable with the data I reported for 1987, but it is worth pointing out that once again there is a considerable turnover in the list of frequently cited sources. Only Nation, Meara, and Richards appear in both the 1987 and 1988 lists of highly cited sources. Sinclair, Channell, Krashen, Renouf, Carter, Halliday, McCarthy, Ostyn and West are all new entrants to the 1988 list of significant sources. With the exception of West and Krashen, all these new sources are part of the Nottingham~Birmingham group of researchers located in the UK. In contrast, Faerch, Kasper, Kellerman, Levenston, Hartmann, Blum-Kulka, Carroll, Haastrup and Lockhart, who were all significant sources in 1987, have dropped out of the highly cited sources list, though all of them continue to be cited to some extent in the 1988 data set.

The 1988 data set is actually quite difficult to work with. Conventional practice is that we base our bibliometric maps on the co-citations among the 100 most cited authors in a data set. With this data set, applying a cut-off of four citations leaves us with 82 authors, while a cut-off of three citations gives us 138 authors, where neither of thee figures is close to 100. Our analysis of the 1987 data used four citations as an inclusion threshold, and so for the sake of comparison with earlier reports, I have adopted the same inclusion threshold for the 1988 dataset. This decision means that authors need to be cited in just over 4% of the papers published in 1988 to be included in our analysis. The Significant Influences are cited much more than this, of course: Sinclair, for example is cited in 18% of all the 1988 papers.

The citation data for the 82 most cited sources were analysed using the Gephi software package (Bastian, Heymann and Jacomy: 2009), and the results of this analysis are reported in Figure 1. Gephi's analysis identifies six clusters, based on how often the members of the cluster are co-cited in the data set.
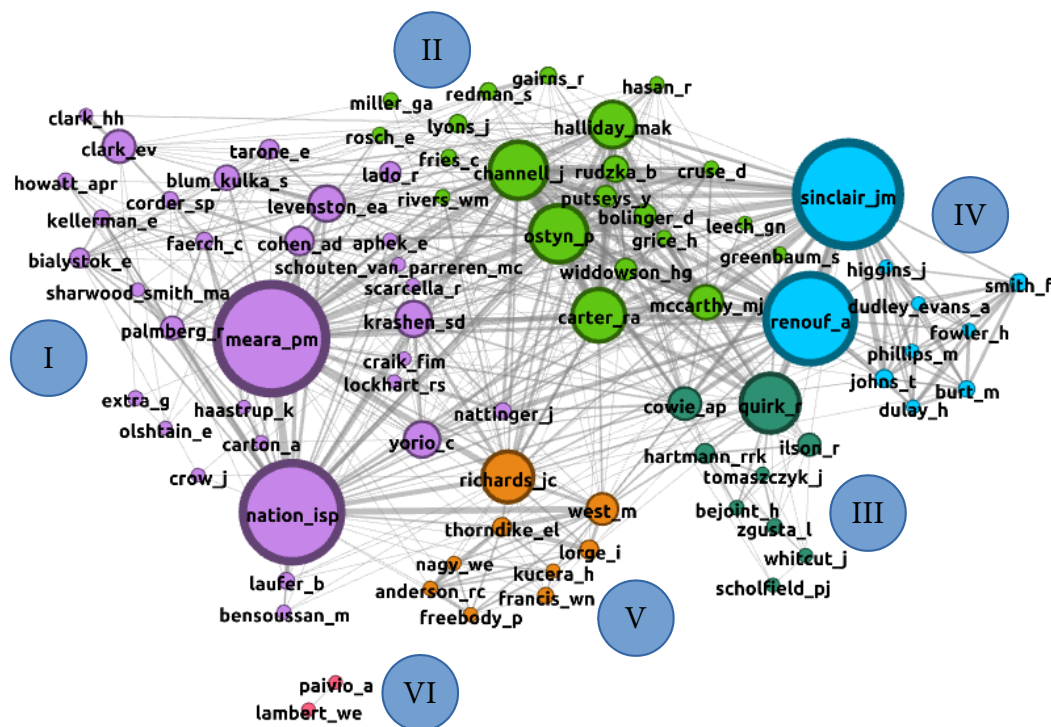
Figure 1: The main clusters in the 1988 data set. The weakest links have been excluded. Nodes are sized according to their betweenness centrality.

Cluster I, at the western edge of the map, dominated by Meara and Nation, is the main vocabulary acquisition cluster. This cluster contains about a third of the sources for 1988, including three of the prolific authors identified earlier. The cluster seems to be mainly concerned with the lexical performance of non-proficient speakers, whether these are bilinguals, L2 learners, or children learning their L1.

Cluster II in the centre of the map, dominated by Channell, Ostyn and Carter, seems to be mainly concerned with meaning, and how semantics can be exploited for language learning. The key source in this cluster is the set of textbooks produced by Rudzka, Channell, Putseys and Ostyn (1981, 1985). These texts, which emphasised an approach to vocabulary based on a componential semantic analysis were identified as important in our earlier papers, and they continue to be influential in 1988.

Cluster III, at the south-eastern sector of the map, is a dictionary research cluster. A cluster of this sort first emerged in the 1987 data set. Here, this cluster appears to be consolidating and reaffirming its position in the map.

Cluster IV, at the Eastern edge the map**,** dominated by Sinclair and Renouf, is largely composed of researchers working at the University of Birmingham. It reflects the growing importance of corpus linguistics in vocabulary research.

Cluster V, at the southern central part of the map, dominated by J. C. Richards, identifies a number of important English word frequency counts, and research that is informed by this work. A cluster of this sort has appeared in all of our maps so far. Anderson, Nagy and Freebody are mainly concerned with how L1 readers acquire the meanings of words that they encounter while reading.

Finally, **Cluster VI**, the small two person cluster at the foot of the map, seems to be the remains of the psycholinguistics clusters which dominated the research earlier in the decade.

Our earlier analyses suggested that this theme has been becoming steadily less important to L2 vocabulary acquisition research throughout the 1980s, and the isolated position of this cluster in the 1988 map seems to confirm this assessment.

In broad terms, there are three fault lines in this map. The main fault line separates Cluster I and Cluster VI from all the other clusters. The sources that make up this grouping are mainly concerned with experimental data, whereas the other clusters might be described as more concerned with descriptive data, and more specifically concerned with descriptive analyses of English. A second fault line separates cluster III and Cluster IV from the other clusters. The sources in this grouping are mainly descriptive linguists, with specific interests in Corpora and Dictionaries. It is slightly surprising that the dictionary researchers in Cluster V are not more closely associated with the corpus linguists. The third fault line seems to lie between Clusters II and III and the rest of the map. The emphasis in this group is on pedagogy, and how linguistic tools can improve L2 learners' experience.

In structural terms, the 1988 map is rather different from the maps that appeared in our earlier reports. The 1988 map has strong interconnections between most of its clusters, and these connections are less dependent on a few key figures who provide the links between clusters. There are two exceptions to this generalisation. The dictionary cluster, cluster III, is not strongly linked to the rest of the network: most of the sources in this cluster are co-cited with each other, but only rarely with other researchers. The same point could also be made for cluster IV. Again here we have a large group of sources who are cited together, but are only rarely cited with sources outside their own cluster. The exceptions here are Sinclair and Renouf, who are frequently co-cited with sources in other clusters, particularly Cluster II.

Probably the most significant feature of the 1988 map is the almost complete absence of any psycholinguistic sources. In 1987 almost a third of the sources fell into two large clusters that we identified as psychologists whose work had influenced research into L2 vocabulary acquisition. Cluster VI is all that is left of this strand of research in 1988. This does not mean that psychological research has stopped, of course, but the analysis does suggest that it has stopped being influential in the work done by applied linguists. This has been something of a trend throughout the 1980s, but it is surprising to see such a rapid collapse. A small number of psychologists do appear in the map, but their influence appears to be limited. Craik and Lockhart are co-cited with Krashen, but not with other members of cluster I. George Miller and Eleanor Rosch, who were previously loosely attached to the main L2 vocabulary cluster, now appear in cluster II. This change in emphasis – a general loosening of ties between two research traditions – feels like a significant shift in the way the field is structured.

We can see the extent of this shift in Figure 2 and Figure 3. Figure 2 shows the new entries in the 1988 co-citation map. There are 42 new entries in this map, slightly more than half of all the sources that appear in figure 1. But more importantly, we are not dealing with a few isolated new entrants here. Rather, the map shows two well-established research areas – one dominated by Sinclair and Renouf, the other by Carter, McCarthy and Halliday – that have quickly become a core part of the L2 vocabulary research. Methodologically, these two areas rely on formal linguistic analysis, and they seem to represent a return to a more traditional relationship between L2 vocabulary acquisition and linguistic theory (cf. Corder 1973). Both themes have

been present in our earlier maps. Nonetheless, the way these two clusters seem to have completely replaced the psycholinguistics clusters that we found in our earlier maps is striking.
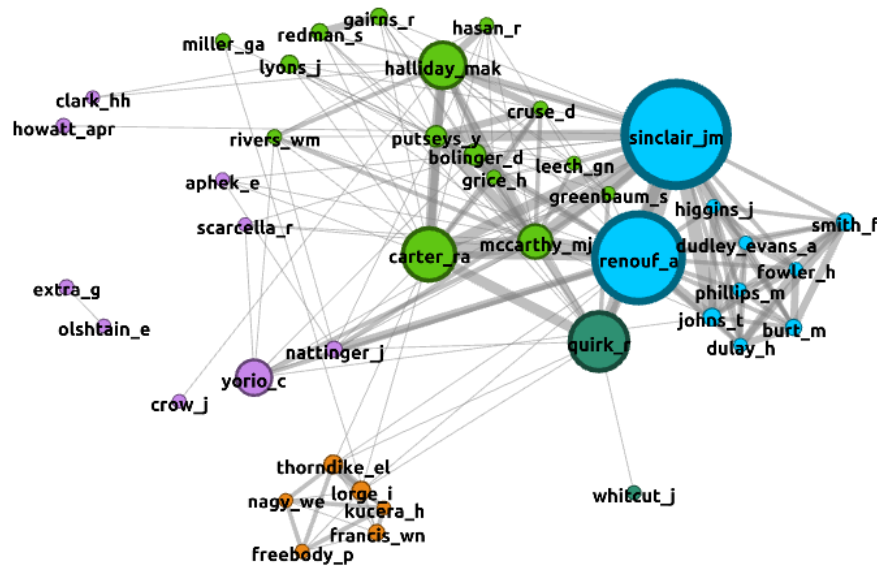

Figure 2: *New entries in the 1988 Co-Citation map*

Figure 3 shows a map of the sources who appear in both the 1987 map and the 1988 map – the "survivors" map. As usual, this map needs to be treated with some caution, as the 1988 map contains more nodes than the 1987 map. Nonetheless, the number of survivors in 1988 is strikingly small: some 60% of the sources in the 1987 map fail to make it into 1988. The majority of the survivor sources belong to cluster I, with Meara and Nation both consolidating their central position within this group. Krashen remains a surprisingly influential source in the cluster. The dictionary research group is remarkably stable: almost all of its members appear in both the 1987 and the 1988 maps, but as we have already noted, this group appears to be somewhat isolated from the mainstream of L2 vocabulary research, and this may be problematic in future years.
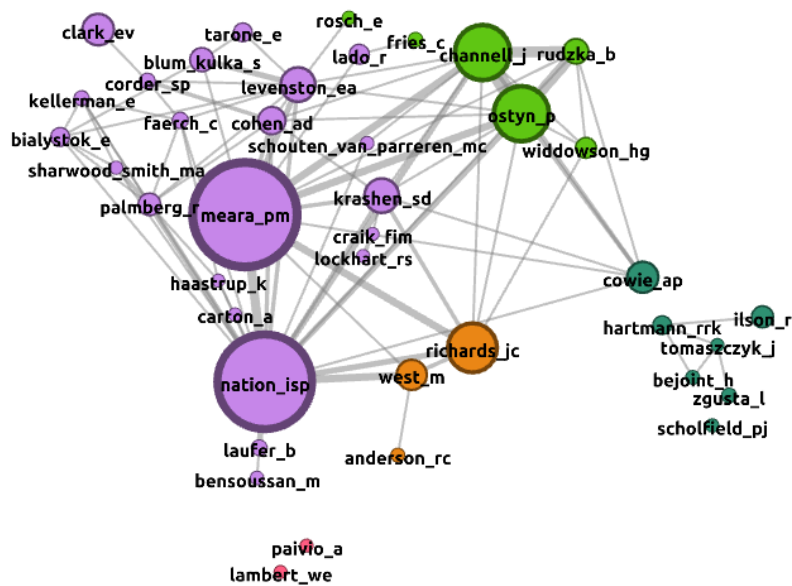

Figure 3: *Sources that appear in both the 1987 and the 1988 co-citation maps*

3. Part 2. A wider perspective: 1984–1988

We now need to look at the 1988 data in a larger context. In our earlier analysis, we were able to do this by looking at a rolling five-year window. This wider context smooths out some of the fluctuations in the annual reports – people who publish a lot in one year but not in adjacent years, for example – and it allows us to identify longer-term trends in the data. In this paper, our five-year window covers the period 1984–1988.

Table 5 recapitulates the main characteristics of the 1983–1987 window which were discussed in Meara (2019). The raw statistics for the 1984–88 data set are broadly in line with the earlier figures, but generally show an increase over the 1983–87 data. The main features of this new data set are reported in Table 6. The data for 1984, 1985, 1986 and 1987 are included in both data sets, but in the current analysis, the 1983 data has been replaced by the larger, and more coherent data set published in 1988. The number of papers included in the new data set is 403, an increase of some 13%, despite the relatively low number of outputs in 1988. As usual, I have not listed all the included papers here, but interested readers can access the list via the VARGA database:(https://www.lognostics.co.uk/varga/). Set the search start date to 1984 and set the finish date to 1988. Then enter ## as the search term, and the program will return a complete list of all the papers included in this data set.

Table 5: *The main characteristics of the 1983–1987 data set.*

Number of papers in the data set:	355
Number of authors contributing to the data set:	326
Number of sources cited in the data set:	3816
Inclusion threshold for this data set	10 citations
Number of cited sources meeting the inclusion threshold	93
Number of cited sources meeting the inclusion threshold	5+2
I: overviews, lexical inferencing, transfer, lexical errors	
II: word recognition in an L2, performance of bilingual speaker	
III: word counts and dictionary use	
IV: semantics and meaning	
V: European vocabulary research (Netherlands and France)	
VI: two disconnected singletons (RC Anderson and Galisson)	

Table 6: *The main characteristics of the 1984–88 data set.*

Number of papers in the data set:	403
Number of authors contributing to the data set:	375
Number of sources cited in the data set:	4080
Inclusion threshold for this data set	12 citations
Number of cited sources meeting the inclusion threshold	100
Number of cited sources meeting the inclusion threshold	7+3
I: vocabulary acquisition (30+1)	
II: word recognition in an L2, performance of bilingual speakers (26)	
III: meaning, corpus analysis (22)	
IV: word frequency counts (5)	
V: Français fondamental (4+1)	
VI: dictionary research (4+1)	
VII: Dutch research (3)	

375 unique authors contribute to these papers, and increase of 15% on the 1983–87 data set. Table 7 shows the number of authors who make N contributions to the data set. As usual, there is a heavy predominance of authors who make only a single contribution to the data set: 77% of the authors fall into this category – almost identical to the 1983–87 figure.

Table 7: The number of authors contributing to N papers in the 1984–88 data set, and the expected number of authors based on Lotka's Law

Papers	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Authors	1							1	1	1	1	5	5	19	53	287
Lotka:	1	1	1	1	1	1	3	4	4	6	8	11	18	32	72	287

Ten authors contributed to five or more papers. Meara contributed to 16 papers; Palmberg contributed to nine papers; Laufer, to eight. Zimmermann and Broeder contributed to seven papers, Carter to six. Arnaud, Beheydt, Extra, McCarthy and van Hout each contributed to six papers in the data set. Three of these authors – Carter, McCarthy and Arnaud – are new to the prolific authors list.

The bottom line of Table 7 shows the number of contributions we would expect to find compared with the predictions made by Lotka's Law (Lotka 1926). Lotka's model (See Appendix 1) suggests that we ought to have many more authors contributing two, three, four or more papers to the data set than we in fact get. The 1983–88 data does look like a power law distribution, but it deviates substantially from Lotka's model when we count the number of authors who produce multiple papers. (Technically, the best fit for this data is has a larger exponent than Lotka's model predicts. The best-fitting curve for the data in Table 7 has an exponent of 2.49. This is a slight improvement on the equivalent figure for the 1983–87 window (2.55). It perhaps indicates that the field is slowly becoming more normalised, though at this stage in its evolution, the field still has a serious over-reliance on authors who contribute to only a single paper in the data set.)

Of course, the fact that some authors contribute several works to a data set does not necessarily mean that their work is influential, so we turn next to the citation data found in the 1984–1988 data set. A total of 4080 sources are cited in this data set, a significant increase on the figure of 3816 that we reported for the 1983–87 window. The data is summarised in Table 8, which shows the number of times the sources are cited in the data set. 65% of the sources are cited in only one paper.

Table 8: The number of times sources are cited in the 1988 data set.

frequency										51	50	49	48	47	46
cases										1					
frequency	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
cases			1			1		1		1			1		3
frequency	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
cases	1		1	1		5	2		4	1	3	4	12	5	6
frequency	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	10	13	9	14	19	18	31	34	34	49	106	151	252	646	2640

At the other end of the scale, a handful of sources are substantially cited. Meara is cited in 51 papers, Krashen in 43, Richards in 40. Levenston is cited in 38 papers, Nation in 36, and Kellerman in 33. Corder, Faerch and Lambert are all cited in 31 papers, Cohen in 30, Eve Clarke in 28 and Channell in 27. This list of significant influences has changed little from the 1983–87 data: Michael West has dropped out of the list; Joanna Channell is a new entry. Meara, Krashen and Richards are all substantially more cited in the 1984–88 data set than they were in 1983–87. Levenston, Nation, Kellerman, Faerch and Cohen show smaller increases in their citation counts. Lambert's count has increased by 1; Corder's count does not change. The number of sources cited at least 10 times in the data set is 138 – almost a 50% increase over the 1983–87 figure (93).

Custom and practice is that co-citation analysis works with the 100 most-cited sources. For our data set, 101 sources are cited at least 12 times, so the analysis that follows is based on this subset of the data. It is worth noting, however, that only 65 of the 1983–87 sources would have met this threshold. It is also worth noting that restricting the analysis to the 101 most cited sources means that we are ignoring a lot of activity which fails to meet the arbitrary threshold. Figure 4 shows the basic map for 1984–88.
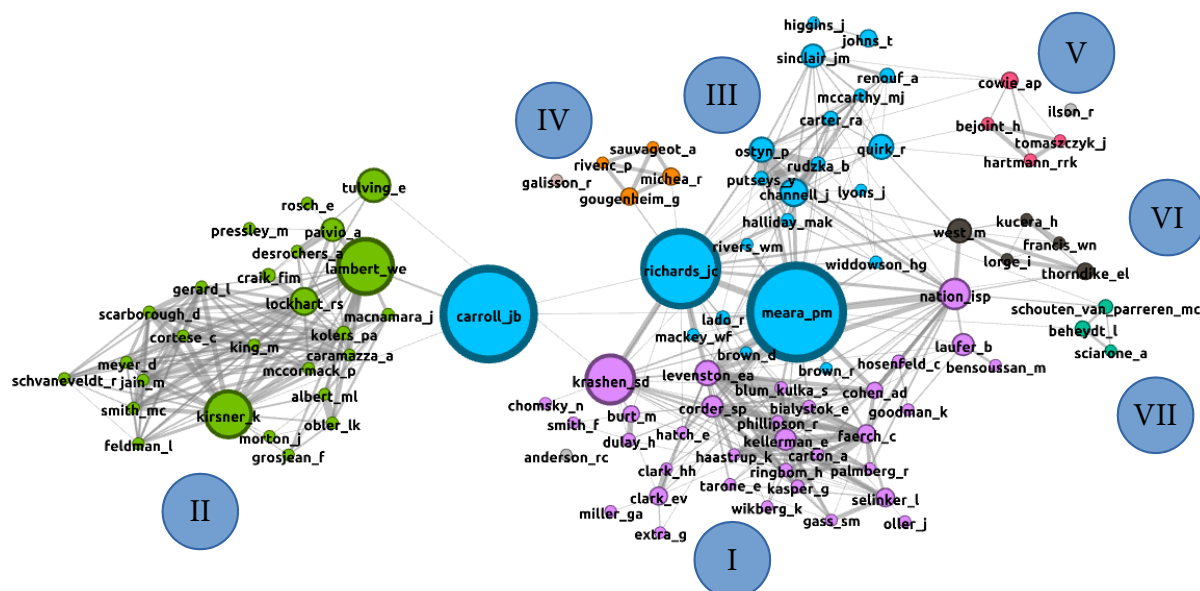

Figure 4: *Patterns of citation among the 101 most cited sources in the 1984–88 data set. Threshold for inclusion is 12 citations in the data set with a minimum co-citation strength of six. Nodes are sized according to their betweenness centrality value.*

Gephi identifies seven clusters in the data set, along with three singletons (Galisson, Ilson and Miller) frequently cited, but not strongly co-cited with other sources who appear in the map. These clusters are broadly in line with the clusters identified in our analysis of the 1983–87 data, but the addition of the 1988 citation data has introduced some changes of emphasis.

Cluster I, the large dense cluster at the bottom of the map, can clearly be identified both in 1983–87 and 1984–88. This cluster contains most of the empirical work on L2 vocabulary acquisition, with a particularly large sub-group of Scandinavian researchers. Several sub-themes can be identified within this cluster: a sub-cluster centred on Krashen, a lexical errors sub-cluster centred on Corder, a reading sub-cluster focussed on Nation and Laufer, a set of L1

acquisition sources (Eve Clark), and a sub-cluster dealing with transfer issues (Kellerman, Ringbom).

Cluster II is the familiar psycholinguistics group of sources at the western edge of the map. This cluster continues to be dominated by Lambert, and the cluster as a whole is very densely interconnected, though it has hardly any immediate connections with the other clusters in this map.

Again, there are a number of identifiable sub-clusters here: Lambert is a key figure in the performance of bilinguals; Kirsner leads a group of researchers looking at word recognition in bilinguals; Craik, Lockhart, Paivio, Desrochers, Pressley and Tulving are all writing about aspects of memory. Rosch, who previously appeared in an L2 semantics and meaning cluster, is a key source in the psychology of meaning. This cluster is about the same size as the equivalent cluster in the 1983–87 map, but it seems to be less densely connected in 1984–88. Our analysis of the 1988 data suggests that few of these sources are consistently cited in the 1988 research, and this suggests that cluster II is likely to shrivel in future analyses.

The main change in the 1984–88 map is the emergence of a strong new cluster apparently centred on Richards, Carroll and Meara. In fact, this cluster, **Cluster III** at the top centre of the map, is mainly composed of new entrants, with Richards, Carroll and Meara appearing to take a leading role mainly because of their high betweenness centrality scores (see below). The key figure in this new cluster is actually Sinclair. The main theme in this cluster is Corpus Linguistics and its application to vocabulary teaching. The cluster also seems to have absorbed an earlier cluster that dealt with meaning and vocabulary acquisition.

That leaves us with four small clusters which formerly appeared as sub-clusters in a larger L2 vocabulary cluster. These are:

Cluster IV (Gougenheim, Michea, Rivenc and Sauvageot) is a group of French researchers, particularly associated with the work on *Français Fondamental*. I have also included Galisson in this cluster, though his work is highly critical of the *Français Fondamental* approach on the grounds that it ignores the cultural and ethnographic aspects of lexical knowledge.

Cluster V (Cowie, Bejoint, Tomaszczyk and Hartmann) is an L2 dictionary group, that previously appeared as a dictionary and frequency count cluster. In this map, the frequency counts emerge as a separate cluster, **Cluster VI** (West, Kucera & Francis and Thorndike & Lorge).

Cluster VII is a small group consisting of Behydt, Schouten-van Parreren and Sciarone. This cluster is basically a Dutch language research group.

There are very few changes between this map and the equivalent map based on the 1983–87 data set. The main clusters remain largely intact in both maps, though there are some small variations where new clusters have budded off from the main cluster and become more independent. In general, these new clusters are characterised by their having very few co-citation links with the other clusters in the map. Sources in Cluster IV for example, are occasionally co-cited with Richards, but they are not co-cited with the sources in cluster VI, despite their overlapping interests. Similarly, the dictionary researchers are sometimes co-cited with sources in cluster II, but they do not figure strongly in the co-citations with Cluster I. The divorce between cluster II (the psycholinguistics cluster) and the rest of the map has become very clear by 1988.

The main characteristic of the new map is that it is extraordinarily stable, despite the changes which have appeared on a year by year basis. We can see this in Figure 5, which shows the "survivor" sources who appear in both the 1983–87 map and the 1984–88 map.
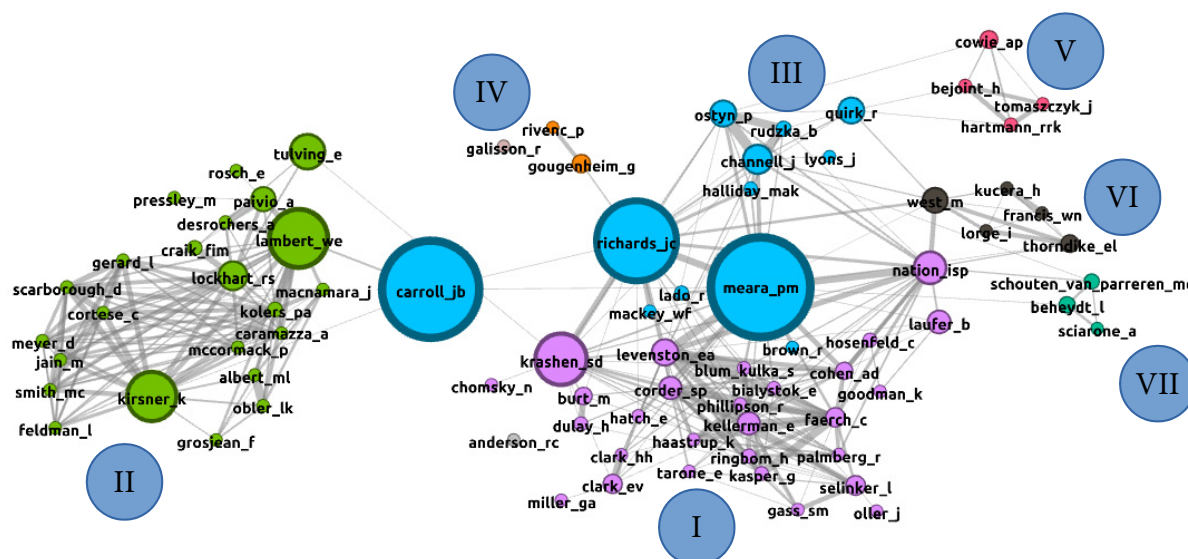

Figure 5: The "survivors". *Sources who appear in both the 1983–87 data set and the 1984–88 data set.*

Figure 5 clearly shows that the number of survivor sources is surprisingly large. In fact, 80% of the sources in figure 5 also appeared in the equivalent map for 1983–87.

The 21 new entrants are shown in Figure 6. Most of these new entrants are single additions to existing groups. Cluster VI and Cluster VII have no new additions. The outstanding feature in this map is the very large new cluster focussed on Sinclair. In thematic terms, this group is a corpus linguistics cluster, concerned with the way corpora can be exploited in language teaching. However, it probably makes more sense to see this cluster as a UK-based geographical cluster. Two UK Universities contribute most of the new sources: Carter and McCarthy were both based at Nottingham University, while Sinclair, Renouf, Higgins and Johns all worked at Birmingham University. For those unfamiliar with UK geography, the distance between these two centres is a mere 80 kilometers, and several of the sources who make up this new cluster worked in both centres. Joanna Channell, a "survivor" who appeared as a significant source in this year's map, was also associated with both centres during this period. Unlike some of the new developments we have seen in previous years, this one looks unlikely to peter out quickly, and we can probably expect a significant growth in this area in future years.
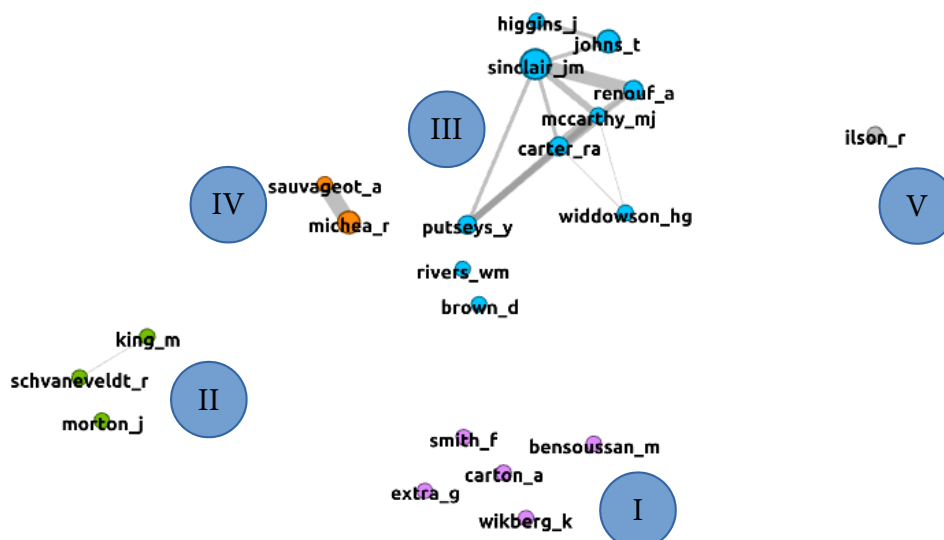
Figure 6: *The new entrants in the 1984–88 map.*

4. Discussion

A number of interesting ideas emerge from these analyses. The first idea is that the overall picture which emerges is much more stable than the pictures we have reported for previous years. Nonetheless, some changes to the field are evident, and the strong emergence of the new corpus linguistics cluster shows this very clearly. What is surprising about this new group is that it does not appear to be strongly co-cited with the other L2 vocabulary sources. We might have expected the corpus linguistics cluster to be strongly connected with the earlier word frequency count cluster, and with the dictionary use cluster, and though there are some links here, they do not appear to be very strong. If anything, the dictionary use cluster seems to have become more detached from the main L2 vocabulary cluster than before.

The second idea is that the research as a whole continues to be overwhelmingly focussed on English research. A couple of small clusters comprising French and Dutch-speaking sources are in evidence, but by 1988 these clusters too have become more detached from the main L2 vocabulary research cluster. I think the main reason for this is simply that none of the sources in these clusters publish very much in English. In contrast, the main L2 vocabulary research does include a very large group of Scandinavian sources (Haastrup, Palmberg, Ringbom, Faerch, Kasper, Phillipson) who do publish in English. Collectively these sources make up a significant proportion of the Cluster I, and their emphasis on lexical inferencing is an important research theme at this time. The danger, of course, is that other researchers who also work in this area, but do not publish much in English (Schouten-van Parreren, for instance) are likely to get overlooked and squeezed out. The small *Français Fondamental* cluster (Cluster IV) looks as though it might be bucking this trend, but this cluster is really an internal dispute among French researchers about the value of the much earlier research on *disponibilité*. English language researchers (with the exception of Richards) seem to have missed the significance of this work.

The small cluster of German researchers which we noted in last year's report fails to make a mark on the 1984–88 map, despite the relatively large number of papers in German which

contribute to the 1984–88 data set. This is partly due to these papers following a different tradition of citation practice from the practices which are by this time becoming normal in the English language research. More importantly, perhaps, while the German researchers cite the main English language sources, English researchers, on the whole, cite each other frequently, but only rarely cite German sources. A good example of this problem is Zimmermann, who was identified in the 1983–87 data set as a prolific author with six publications. A further publication in 1988 brings his total to seven – more than enough for him to retain his place in the list of prolific authors. Altogether, though, these seven publications garner only ten citations in the data set (and seven of these are self-citations). This means that Zimmermann fails to appear in the 1984–88 map in spite of being a prolific author, and the related German research disappears with him. Clearly, publishing a lot of papers does not automatically mean that they will be widely cited.

The third feature which emerges from the 1984–88 map concerns the role played by a few key figures in the structuring of the map. We noted in last year's report that the key figure in the 1983–87 map was Krashen, with Richards, Meara and Lambert playing lesser roles. This judgement was based on a measure called **betweenness centrality.** The theory behind the betweenness centrality measure is that some nodes in a map are key to its structure because they provide links between the different clusters, and thus represent important points of contact between different research groups. The measure is based on the probability of a node being found along a path that links randomly chosen pairs of nodes. Nodes which appear frequently in these random paths score highly on the betweenness centrality measure, and when the clusters are highly divergent, the few nodes that link them score very highly on this feature. In practice, most sources tend to be co-cited with other sources in their immediate cluster, and only a few sources are co-cited with sources from two or more clusters.

In the 1984–88 map, there has been a significant shift in the make-up of these key nodes. Lambert continues to dominate cluster II, but is not generally cited in the broader L2 vocabulary acquisition literature. Krashen's influence has not entirely disappeared, but it is much reduced. Richards and Meara have become much more influential in the 1984–88 map: both are highly co-cited with the members of the new corpus linguistics cluster, and Richards' early work on *Français Fondamental* provides a strong co-citation link between Cluster II and the French researchers in Cluster IV. The surprise feature in the 1984–88 map is the importance of J. B. Carroll. In the 1983–87 map, Carroll is closely associated with a cluster that is focussed on word counts and dictionaries, but in the new 1984–88 map, he appears as the key link between Cluster II – the psycholinguists – and the rest of the network. My immediate reaction to this was that the analysis was over-estimating Carroll's real influence on the L2 vocabulary research, since, unlike the other key figures, he does not actually publish on L2 vocabulary acquisition during the period we are analyzing here. On reflection, however, I think this reaction was wrong. Twenty-two of the papers in the 1984–88 data set cite Carroll's work, but these citations are not limited to a single influential paper, or a paper that accidentally bridges the gap between the linguistic and the psychological approaches to L2 vocabulary. Most of the citations relate to Carroll's word frequency count work (Carroll, Davies and Richman 1971), or his Age of Acquisition norms (Carroll and White: 1973a, 1973b), but the citations are not limited to these useful tools. Particularly important are a 1964 paper published in the *Harvard Educational*

Review (Carroll 1964a) and the monograph *Language and Thought* published in the same year. (Carroll 1964b). In this case, the co-citation approach does appear to have successfully identified a significant influence who might otherwise have been overlooked.

Finally, it is worth noting that the 1984–88 map highlights the continuing decline in the influence of psycholinguistic research on L2 vocabulary research. Although Cluster II has seen the arrival of three new sources in 1988 (Morton, King and Schwaneveldt), it does not show any real growth in this time window. This cluster is slowly crystallising into three sub-themes (bilingual word recognition, the skills of bilingual speakers, and some residual work dealing with imagery and depth of processing). The rich connections in this cluster are at least partly due to differences in authorship practices in psychology and linguistics. Papers published in psychological journals are more likely to have multiple authors than are papers in linguistics. This results in dense clusters of co-citations, which make this work look more important than it might really be. The important links in a co-citation analysis are those which strongly link the different clusters: the 1984–88 map strongly suggests that cross-cluster links of this sort are in short supply.

5. Conclusion

The main point to emerge from the analysis presented in this paper is that L2 vocabulary research between 1984 and 1988 is remarkably stable, though we still find significant changes on a year by year basis. New research areas have added a degree of focus to the research published in 1988 year, and we have seen that the research is becoming more obviously dependent on linguistics for its main points of reference, whereas the research in earlier years cited a more eclectic set of sources. Some of the Significant Influences who played pivotal roles in our earlier analyses are beginning to be much less influential.

The next paper in this series will examine the research published in 1989 in the context of a five-year window covering 1985–89. Will the field continue to develop in the same way, or will we see significant shifts of direction in this period? It is hard to tell. All we can say at the moment is that by 1988 something that resembles a recognisable vocabulary research agenda is beginning to emerge.

References

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Carroll, J. B. (1964a) Words, meaning and concepts. *Harvard Educational Review* 34(2): 178–202.

Carroll, J. B. (1964b) *Language and thought.* New York: Prentice-Hall.

Carroll, J. B., P. Davies and B. Richman. (1972). *The American heritage word frequency book.* Boston: Houghton Mifflin.

Carroll, J. B. and M. N. White (1973a). Age-of-acquisition norms for 220 picturable nouns. J*ournal of Verbal Learning and Verbal Behavior* 129(5): 563–576.

Carroll, J. B. and M. N. White (1973b).Word frequency and age of acquisition as determiners of picture naming latency. *Quarterly Journal of Experimental Psychology* 25(1): 85–95.

Corder, S. P.1973. *Introducing applied linguistics.* Harmondsworth: Penguin Education. 1973.

da Solla Price, D. (1965). Networks of scientific papers. *Science* 49(3683): 510–515.

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. R*ELC 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 and A. Downing (eds.), *The Functional Perspective on Language and Discourse: Applications and implications*, 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. (2018). Laying the foundations. A bibliometric analysis of vocabulary research in 1982–86.*Linguistics Beyond and Within* 4: 108–128.

Meara, P. M. (2019). Steady as she goes? A bibliometric analysis of L2 vocabulary research in 1987. *Linguistics Beyond and Within* 5: 103–119.

Meara, P. M. (no date) *VARGA: The Vocabulary Acquisition Research Group Archive.* Available at: https://www.lognostics.co.uk/varga/ (accessed 22nd October 2020).

Perdue, C. (1982). *Second language acquisition by adult immigrants: A field manual.* Strasbourg: European Science Foundation.

Perdue, C. (ed.) (1993). *Adult language acquisition: cross-linguistic perspective.* Vol. 2. Cambridge: Cambridge University Press.

Rudzka, B., J. Channell, Y. Putseys and P. Ostyn. (1981). *The words you need.* London: Macmillan.

Rudzka, B., J. Channell, Y. Putseys and P. Ostyn. (1985). *More words you need.* London: Macmillan.

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.

White, H. D. and B. C. Griffith.(1981). Author co-citation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* 32: 163–171.

Acknowledgements

Appendix 1: Lotka's model

Lotka (1926) suggested that there might be a straightforward relationship between the number of authors who contribute a single paper to a field and the number of authors who make multiple contributions to the field. Suppose, for example, that we have 250 authors who make a single contribution to a data set, then it would be unusual to find only a single author making two contributions, and it would likewise be very unusual to find that a single author makes twenty contributions, while no other authors make more than one contribution to the data set. Lotka suggested that the expected relationship could be described as a power law:

$$E_N = T / N^x$$

where   T is the total number of authors who contribute a single paper to the data set,
             N indicates 2,3,4,5… outputs,
and E_N is the expected number of authors contributing to N outputs.

In practice, the value of x (the exponent in Lotka's formula) is usually around 2 – that is, a value of 2 for this exponent gives a fair approximation of what happens in real life. So, for a data set in which 250 authors contribute to just one paper in the data set Lotka's model predicts that we can expect $250/2^2 = 63$ authors who contribute to two papers in the data set, $250/3^2 = 28$ authors who contribute three papers to the data set, $250/4^2 = 16$ authors making four contributions to the data set, and so on as shown in the table below.

contributions	10	9	8	7	6	5	4	3	2	1
Expected E_N	2	3	4	5	7	10	16	28	140	250

Clearly, this model predicts that the number of papers an active researcher might be expected to produce falls off rather quickly. Empirical tests of what has become known as "Lotka's Law" do seem to work well. However, the model works best when we are dealing with well-established fields, and very large data sets. The single year data sets that I have discussed in this series of papers are not a close match to Lotka's expectations, but the larger 5-year data sets are generally a better fit to the power law model. In both cases, however, we get a much better fit when the value of N^x is raised above 2. For example, we get the best fit for the 1988 data when x= 5.1, though this figure needs to be treated with some caution because the data set is relatively small. Higher values of x seem to be typical of immature, highly volatile fields. Generally speaking, the exponent values we find for the L2 vocabulary research literature are higher than we would normally expect. I do not yet fully understand the implications of this.

Appendix 2: Co-citation analysis: The methodology

The co-citation method used in this paper 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 da Solla Price (1965), has been extensively used to analyse research in the natural sciences (e.g. White and 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 and 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.

A dictionary of legal English collocations as an aid for mastering the legal English genre

Katarzyna Mroczyńska
Siedlce University of Natural Sciences and Humanities, Poland

Abstract

The article begins with a brief overview of collocations and their features as being central to legal language and, as such, worth studying, especially in view of the fact that legal language studies tend to be mostly interested in terminology rather than phraseology. To bridge this gap, the article offers a tool for legal English learners, i.e. the dictionary of legal English collocations based on judgments of the UK Supreme Court. Our dictionary project is aimed at analysing the corpus we created, using Sketch Engine software, a cutting-edge lexicographic tool which enables the uploading and exploration of users' own corpora. The project will focus on analysing bipartite legal English collocations appearing in the corpus. The next stage of the project will be the preparation of the final product of our research, i.e. a dictionary of legal English collocations. We believe that such a dictionary will prove a useful aid for mastering the conceptual structure of legal English.

Keywords: collocation, dictionary compiling, legal English

1. Introduction

In this article, we intend to present our ongoing project which involves compiling a dictionary of legal English collocations. This article opens with a brief summary of research findings regarding collocations, where we advocate the need for mastering collocations in order to become a proficient user of legal English. Based our observations as academic teachers working with students of Business English, including elements of legal English, we notice that students often lack sufficient exposure to phraseology which forms part and parcel of the language and consequently, they frequently struggle to find appropriate collocations in a given context. In an attempt to remedy this situation at least partly, we offer a tool for legal English learners, i.e. a dictionary of legal collocations based on judgements of the UK Supreme Court published in the period from 2009 to 2018. In the functional classification of legal texts, judicial texts belong to the group of hybrid texts, i.e. they are primarily descriptive but also prescriptive, as opposed to the other two types in the classification offered by Šarčević (1997: 11), which are purely descriptive or purely prescriptive. We believe that the Supreme Court judgements represent high-quality legal documents and, as such, may be treated as a good point of departure for a

reliable analysis of the legal English genre and may equally serve as a rich source of textual material to be used in teaching/learning. Our dictionary project[1] is aimed at analysing the corpus we created using Sketch Engine software, a cutting-edge lexicographic tool which enables uploading and exploring users' own corpora. Section 4 contains a more detailed description of functions the software offers and of the statistical measures it uses.

As legal language studies appear to concentrate mostly on terminology rather than phraseology, we would like to try and bridge this gap. To this end, the project will focus on analysing bipartite legal English collocations appearing in the corpus with the dictionary of legal English collocations, which we hope will have some pedagogic applications, to follow as the final stage of the project. Given that professional discourses such as law include a large number of formulaic expressions, multiword units and collocations, the lexical items to be presented in the dictionary will be worth integrating into a teaching curriculum so that students are exposed to these patterns, explicitly to enhance their learning outcomes.

2. Collocation explained – definitions and competing approaches

Since terminology makes an attempt at analysing, recording and describing the concepts of a specialised language, it may be of great help in providing a better and more thorough understanding of legal concepts (Bajcić 2017: 7). We tend to believe that the shift in trends regarding the development of terminology may prove useful in the field of law, which is of key interest to us. The field of law relies on the language to express legal concepts so as to produce a desired effect. However, it should be noted that the discipline of law is based on concepts which frame the professional knowledge in this area rather than words. Moreover, legal concepts need to be analysed within their extralinguistic context, as this context is of paramount importance, providing details necessary to come up with a clear and reliable interpretation of legal issues.

As Biel (2014: 42) claims, the conceptual structure and the systemic nature of legal terminology affects legal phraseology. Therefore, phraseology, and specifically collocations, act as a tool to express complex interrelations between legal concepts. Moreover, it is also capable of entering into relations with other terms. Thus, we decided to put collocations in the centre of our lexicographic project devoted to the legal English genre.

In the literature dealing with lexicological theory and lexicographic practice, we may encounter numerous definitions of the term *collocation*. To make things more complex, it seems that there is no universally accepted formal definition of collocations, not even a proposal for the definition (Mel'čuk 1998: 23).

The definitions will vary depending on the approach they have been based on. Among the three most popular approaches applied by linguists researching collocations are a frequency-

---

[1] Devising a dictionary of legal English is a project run at the Siedlce University of Natural Sciences and Humanities by Katarzyna Mroczyńska, the author of this article, and Tomasz Michta, PhD. The researchers are Assistant Professors at the Department of English Studies and Translation Studies of the Institute of Language and Literary Studies.

based approach, a semantic-oriented approach and, last but not least, a relatively new, pragmatically-driven view (Siepmann 2005: 410).

The first approach, advocated, among others, by Mel'čuk (1998), Gonzalez-Ray (2002) and Hausmann (1997), assumes the existence of a particular meaning relationship between the constituents of a collocation. As an advocate of this approach, Mel'čuk (1998: 23–24) argues that collocations form a subclass of *set phrases*, or *phrasemes*, understood as lexical units characterised by their non-compositionality; phrasemes need to be treated, stored and used as a whole. Moreover, Mel'čuk (1998: 26), in his formalist proposal, introduces two ancillary concepts for distinguishing phrasemes from free phrases, i.e. the concept of unrestictedness ("unlimited freedom of choice among equivalent independent meanings and expressions") and of regularity ("observance of general rules in combination of meanings and expressions and is related to the concept of combination rules of language"). Accordingly, the term *phraseme* will refer to linguistic signs including a signified and a signifier which can be constructed neither unrestrictedly nor regularly.

Mel'čuk (1998: 28–29) also devises a formal definition of a collocation. He claims that collocation AB is a phrase whose signified includes the signified of one of its two constituent lexemes, e.g. A, which is freely chosen by the speaker, but the other component, a signified C, is such that the lexeme B expresses C, and it is chosen contingent on A, which means that the signifier of a collocation is not unrestrictedly constructed.

This definition may be expressed in the following formula (Mel'čuk 1998: 28):

AB = ⟨'A⊕C';/A⊕B⟩|'C' is expressed by **B** such that /A⊕B/is not constructed unrestrictedly.

Collocations, or *semi-phrasemes*, which are the focus of our attention, constitute the majority of the entire phraseme inventory. To describe them accurately and systematically, Mel'čuk (1998: 29–30) uses the Meaning-Text theory and the comprehensive concept of Lexical Functions, whose detailed description is beyond the scope of this study.

In the frequency-based view (represented by Sinclair 2004, or Kjellmer 1994, for example), on the other hand, we will examine collocations understood as a statistically significant co-occurrence of at least two or more words. We can measure the strength of a given collocation, or a syntagmatic attraction, by means of specialised tools and statistical analysis of corpus data. These statistical calculations reveal which word pairs yield significantly more co-occurrences than what would be expected by chance, taking into account the words' total frequencies in the gathered corpus material. In this manner, we are capable of establishing the most significant collocates of any given word in the language genre that the collected data represents (Sinclair 1966: 418; Hoey 1991: 6–7).

According to Lehecka, the collocation strength, i.e. the attraction between a node (W1) and its collocate (W2), is calculated based on four observed absolute frequencies in the data:

(i)   the total number of word tokens in the corpus,
(ii)   the number of tokens of W1 in the corpus,
(iii) the number of tokens of W2 in the corpus, and

    (iv)   the number of tokens where W1 and W2 co-occur within a specified distance (in the collocation window) from each other so that they are regarded as co-occurring (Lehecka 2015: 2).

The next step in the analysis is the comparison of the observed number of co-occurrences in the corpus and the expected number of co-occurrences, i.e. the number expected by chance given (i), (ii) and (iii). If the observed number of co-occurrences of the node and the collocate is larger than what can be ascribed to chance, then W2 is a statistically significant collocate of W1. Individual words in a language have very different frequencies. That is why the collocation strength between different word pairs needs to be done by using a statistical association measure which takes into consideration the uneven distribution of words in the data and not absolute frequencies alone, which would be insufficient for the needs of reliable collocation analysis. Nowadays, statistical tools offer over 50 different association measures (Evert 2009: 1,243), with the z-score, the t-score, MI (Mutual Information), the log likelihood ratio and Fisher's exact test being most frequently applied in the collocation analysis. Undoubtedly, the association measure a researcher chooses has a great impact on the results of the analysis, and as such, this choice certainly requires careful consideration. A thorough discussion of the advantages and the disadvantages of different associations can be found in Evert (2005); or a more concise one in Wiechmann (2008) and Evert (2004) and (2009) (Lehecka 2015).

    The third approach, a younger relative of the two presented above, makes an attempt at explaining structural irregularities and non-compositionality underlying phrasemes and collocations by referring to pragmatic regularities which provide a link between the situational context and the linguistic form, which brings us back to the cognitively-based contextualisation theory offered by Fillmore (1976) (Siepmann 2005: 410–411).

    All three approaches are not as divergent as they may appear at first glance, and some scholars seem to have found some common ground to build a unified view of collocations, at the same time indicating certain drawbacks of each of the individual approaches. The proposition devised by Siepmann (2005 and 2006) may be worth a more detailed presentation, as it offers a new, broader definition of *collocation* and a revised classification of the collocational spectrum.

    Siepmann points out major drawbacks linked to each of the three presented approaches, i.e. the frequency-oriented approach provides us with just the raw data but does not reveal much about how this material was formed or how it should be structured; despite being appealingly intuitive, the semantically-driven view of collocations does not offer a holistic view and will only be fragmentary; and finally, the pragmatic approach, in its pure version, will fail to account for the collocations of semantic features. Therefore, he advocates an extension of the semantically-based approach "that will take account of strings of regular syntactic composition which form a sense unit with a relatively stable meaning" (Siepmann 2005: 411) based on a rigorously carried out study of the linguistic corpus. Incorporating the findings of the cognitivist camp, he also postulates loosening the definition of collocation so that it encompasses the concepts of usage norm and statistical significance, on the one hand, and the holisticity of the collocational unit, on the other hand. Collocations may be considered as self-

contained "wholes", given the fact that native speakers can ascribe meaning to general language collocations even if they lack their context (Siepmannn 2005: 438).

Siepmann (2005: 431) offers the following typology of collocation phenomena, which may cover the entire range of morpho-syntactic constructions:

a) colligation which concerns both grammatical preferences and those of longer syntagms, e.g. *far be it from me to* + INF;

b) collocation between lexemes and phrasemes, e.g. *in the end*;

c) collocation between lexemes and semantic pragmatic feature, e.g. [expectation] = *duly*;

d) collocation between semantic pragmatic feature, or long-distance collocation.

It is worth noting that instead of being based on colligational relations, some collocations are just based on semantic relations.

Although researchers seem to find it difficult to adopt a uniform, widely-accepted definition of collocation encompassing all the linguistic features of the phenomenon, a list of criteria of collocability, which will allow us to classify a given phraseological unit as collocation, can be compiled that will be basically the same across various research on collocation. These criteria include:

a) frequency of co-occurrence;

b) combinatory restrictions;

c) degree of compositionality;

d) degree of transparency;

e) span of words between node and collocate, or collocational window (Patiño 2014: 122–124).

When we adopt such a broad definition of collocation as the one offered by Siepmann, we notice that collocations are extremely widespread in the language, and as such, they seem to dominate the language use. What is more, being structurally complex units, collocations in this sense, and not just individual words, may be treated as linguistic signs on their own (Siepmann 2005: 438).

As collocations seem to dominate the language use, both general and specialist, we decided to undertake a lexicographic project aimed at analysing genre-specific collocations appearing in legal English with the assumption that they will reflect the conceptual structure of the language of the law. We were encouraged by the fact that most legal language research focuses on terminology rather than phraseology (which is relatively unknown ground). Specialist dictionaries and term bases include terms, but they do not customarily contain the collocational relations of these terms, and even if they do, collocations tend to appear in the exemplificatory material, thus making it more difficult for the user to identify them. Indeed, as dictionaries of LSP collocations are almost non-existent, we would like to bridge this gap and compile a dictionary of legal English collocations, which is the planned outcome of our research project. Although Siepmann (2006: 34) argues quite convincingly that a bilingual or multilingual dictionary has obvious advantages over a monolingual approach, its compilation inevitably is more time consuming and, in the case of a culture-bound subject field such as law, entails solving numerous problems of equivalence that stem from the asymmetry between different legal systems, a problem not to be downplayed by any lexicographer. Therefore, we decided to

compile a monolingual dictionary of collocations, hoping that in doing so we will be able to reach a wider audience (an English-Polish dictionary of legal collocations is more likely to appeal only to speakers of both English and Polish) and that our dictionary will be followed by bilingual dictionaries of legal collocations.

It is also worth mentioning that the concept of collocation does not only refer to textual statistics, but it reflects a mental representation of the lexicon, as collocations are formed by the cognitive process of priming. Hoey distinguishes three elementary types of priming: collocation, colligation and semantic preference/association, with the priming of lexical items with collocations in this psychological sense being the foundation of language structure in general (Hoey 2005: 8–9). In light of these findings, we may assume that knowing how words collocate is part of what it means to know a language or a genre of a language. Consequently, a dictionary of collocations may serve as an aid in mastering the conceptual structure of an analysed language genre.

3. Assumptions for the preparation of a dictionary of legal English collocations and applied tools

Dictionaries have accompanied people for a relatively long time, and they have served as an aid to understand a given domain of knowledge. This assistance may be particularly needed in relation to the legal domain, which affects our daily activities throughout our whole lives. Although intended to be accessible, legal regulations frequently turn out not to be understandable to non-legal professionals. It seems that this is not the language of the law as such, but the law itself that is complex (Bajčić 2017: 140). What is more, concepts used in the domain of law may lack clear-cut boundaries, which is seemingly at odds with the need for precision expected on the part of legal text users. Legislators strive for drafting precise, unambiguous rules, which, at the same time, will be able to accommodate the meaning in a way general enough to be applicable to as many situations as possible; combining the necessary precisions and the need for generalisation is one of the challenges law faces (Wagner and Gémar 2013: 179).

When it comes to the purpose of a legal dictionary, we are inclined to adopt the view offered by Bajčić, who asserts that "the main purpose of a legal dictionary is to enable users to learn about legal concepts in order to understand the law" (2017: 138).

Bearing in mind the fact that collocations reflect a language's conceptual structure, and the ability to use collocations in a correct and natural manner represents the user's mastering of the language within a given specific genre, we believe that a dictionary of English legal collocations may contribute significantly to the improvement of knowledge of the language and also of the workings of the law as such. Given that the framing of ideas in a foreign language depends on the user's linguistic competency and that we (as users) communicate by means of semi-prefabricated lexico-grammatical units rather than individual words, the learners or non-native users of a given genre need to learn a set of lexical items. Learning collocations seem to play a key role in the language acquisition process, and the set of collocations which needs to be learnt

for a person to become a competent user may be determined using the criteria of frequency, availability and learnability.[2]

Thus, the aim behind our project is to draw up a comprehensive frequency list of one hundred key legal terms, specifically nouns, and analyse their collocational behaviour. We assume that this will lead to the identification of terms and collocations that have to be mastered to be able to perform at the near-native (or lower) proficiency level.

To ensure the most reliable and up-to-date input representing examples of real-life application of the legal language, we have decided to devise our collocations dictionary on the corpus of judgements made by the UK Supreme Court, which is the final court of appeal for civil, as well as criminal, cases in England, Wales and Northern Ireland and therefore plays an important role in the development of the country's legal regulations. Moreover, the judgements are public, so their wording is published and made available on the Court's website (supremecourt.uk). To create a corpus which, as we believe, will yield reliable results and improve the depth of coverage, we decided to cover the judgements during the ten-year period from the Court's inception in 2009 to 2018. However, we are aware that the corpus material collated in this manner and only comprising the last ten years is narrowly synchronic and may display some deficiencies, e.g. it may not reflect the knowledge and experience of language users accumulated over several generations. Consequently, it is worth noting that a collocation may still exist despite the fact that it is absent from a corpus. Moreover, we are aware of the fact that that a corpus based on one source may not suffice to reflect fully the richness and complexity of the legal genre. Therefore, we see an opportunity to embark on a new more ambitious project which would rely on a fully-fledged corpus based on different types of legal texts and would provide a broader spectrum of legal phraseology.

According to Nielsen (1994: 33), while compiling a dictionary which would cover LSP communication it is of key importance to establish the target group of users and identify their lexicographic needs. To perform this task, the researchers working on the projected dictionary, who are also academic teachers specializing in LSP courses, carried out observations during their work and they noticed that, otherwise proficient language users, students often lack linguistic competence and confidence in the area of phrasemes or collocations. The researchers analysed the information regarding the use of existing LSP and general dictionaries gathered while running LSP and translation courses and it appeared that students were often unable to find a collocation they were looking for in the available dictionary resources. This gap in the resources available to students that may be bridged by the projected dictionary of legal English collocations. The dictionary is a monolingual work based on a limited set of texts and it will be used a pilot project that, if successful, may be extended into a bilingual dictionary based on a wider range of legal texts in the future.

The projected dictionary is intended to be a single-field (legal) dictionary for students of business and legal English, or LSP learner's dictionary. As it is designed to be used by students, it will contain a smaller lemma stock than LSP dictionaries for professional users.[3]

---

[2] See Siepmann (2006: 4–9).
[3] See Nielsen (1994: 41).

Undoubtedly, collocations deserve exhaustive treatment in LSP dictionaries but this need is often difficult to meet in general field dictionaries or multi-field dictionaries for reasons of limited space in the publication. Therefore, we see some room for the projected dictionary, focusing solely on the area of legal English collocations, which makes an attempt at remedying such drawbacks.

While compiling a dictionary, it is also crucial that the genuine function of a dictionary is clearly determined as it is the best criterion for (not) including particular types of information in the dictionary. According to Nielsen (1994: 44), dictionary functions may refer to text translation, composition or comprehension. As the projected dictionary is intended to provide assistance in producing texts (or utterance), its main function will be composition. To make foreign language texts or utterances, the user of a composition dictionary will need examples of language use and such real-life examples culled from the corpus accompany dictionary entries.

The determination of the target user group and the main dictionary function should be reflected in the publication macro- and microstructure. As for the former, the dictionary will have a complex macrostructure as it will contain a preface, user's guide, word list, and a cross-reference list. We decided to include these components in the dictionary as we see them as clearly and directly related to each other and adding necessary information to the scope and function of the dictionary in question. Our intention is to produce a dictionary whose all individual components will constitute a simple and coherent whole: a dictionary of legal English collocations.

We are going to follow the established rules and our dictionary will include the front matter, the word list and the back matter. The front matter will contain among other things the list of contents and the preface which gives authors' explanatory remarks concerning the dictionary function, scope and application as well the corpus on which the dictionary is based. What will also be included is a user's guide which will explain all the functional elements contained in the dictionary so that users do not have to guess themselves the meaning of various abbreviations of functional elements, for example.

As for the word list, we applied a frequency method to compile this list. The word list of the projected dictionary contains around 100 alphabetically arranged entries being legal English nouns with the highest frequency in the analysed corpus. The nouns are selected based on the frequency list generated by the statistical tools offered in SketchEngine software. It is worth noting that with this method of dictionary compilation based on a computerised corpus it is easier to find illustrative examples of the LSP terms usage. Needless to say, such examples are extremely valuable as a pedagogic aid both for students and teachers dealing with the legal English genre.

As for the back matter, the projected dictionary of legal English collocations will contain a cross-reference list giving easy access to all occurrences of a given noun in the dictionary, which as we believe is relevant to the use, scope and purpose of the dictionary.

The lemma list of a legal dictionary may be expected to feature solely specialist terms or words which have at least one specialised meaning within the field of law. Consequently, lexicographers apply the so-called pruning process where general language terms are excluded unless they have a special legal meaning differing from the normal meaning of the analysed terms (Nielsen 1994: 154).

After making some preliminary provision for the macrostructure and the word list for the projected dictionary, we also need to consider its microstructure. For the purposes of this article, we adopt the definition offered by Hausmann and Wiegand which says that "the microstructure of a dictionary article is the total set of linearly ordered information items following the lemma" (1989: 340). One of the conclusions we may draw based on this definition is that the microstructure is an ordering structure, its main function being an optimal arrangement of the article-internal information items. This optimal manner of organisation is important to a target user. Therefore, we intend to implement a microstructure which presents information in an easy to understand way and results in a user-friendly article, coherent and easy to read. The type of microstructure used in a dictionary depends on the nature of the dictionary itself, its purpose, scope, application and also its target group (Nielsen 1994: 223).

For the microstructure of the projected dictionary of legal English collocations, the following microstructure was adopted: a lemma form realisation item, i.e. a spelling form, collocations arranged by type, i.e. premodifier + TERM, TERM + noun, verb + TERM, TERM + verb, preposition + TERM, TERM + preposition, with each collocation section followed by a corpus-based example. Polysemous lemmata will have separate article sections devoted to various meanings each and following the adopted microstructure for a dictionary article.

Having established the target user group and the key function of the projected dictionary as well as having presented its macro- and microstructure, we will go on to discussing the analysis of the gathered corpus material and tools used to carry out this task. In order to facilitate the analysis of the linguistic material, Sketch Engine software,[4] a leading corpus tool widely used in lexicography, was applied. The name of the software comes from one of its functions, i.e. the wordsketch, which is a concise, yet extremely comprehensive, summary of a word's grammatical and collocational behaviour presented in a transparent one-page figure. Actually, the word sketch can be regarded as a draft dictionary entry, as the system has already found, analysed and organised all the recurring patterns for the word so that they are at the lexicographer's fingertips to be edited, reorganised and finally published (Kilgariff *et al.* 2014: 8–10).

Apart from the iconic word sketch mentioned above, the software also includes the following functions (Kilgarriff *et al.* 2014):

1. Concordance – a basic tool for any researcher working with a corpus, as it is capable of searching a corpus for a word form, a lemma, a phrase, a part of a speech tag, etc., going back to the raw data underlying any analysis. The system converts all queries into Corpus Query Language (CQL), which can be used directly.
2. Word List – the software may generate frequency lists of words, lemmas, n-grams or key words.
3. Keywords and Terms – this function enables extraction of core lexis in a corpus using a "keyness score".
4. Collocations – the tool calculates words that are statistically associated with the query term. The system uses several statistically relevant measures to find the best collocation candidates, such as T-score, MI, log likelihood, logDice, etc.

---

[4] See http://www.sketchengine.eu.

6. Word Sketch Difference – this function offers a comparison of two words based on collocations.
7. Thesaurus – creates a distributional thesaurus based on common collocation. The resulting list of words includes items in various semantic relationships.
8. Trends – this function is useful when conducting a diachronic analysis of word usage.
9. WebBootCaT – it is a set of programs that may be used to compile a user's own web corpus.

One of the advantages of the tool is the fact that apart from providing pre-loaded corpora of its own, the software also allows the user to upload, build and explore their own corpus using the WebBootCaT tool. What is more, when it is applied to seed words from a specific domain, this corpus-building procedure seems to be an extremely efficient method of discovering the terminology and phraseology of a specialist domain. Additionally, the software allows a researcher to accurately tokenise, lemmatise and tag the corpus by specific parts of speech. Tools for these processes are available for selected languages, including English (Kilgarriff *et al.* 2014).

For the purpose of analysing the material, we gathered and uploaded corpora comprising the judgments made by the UK Supreme Court in order to create our own corpus for the project. The next step was to use the Word List function to generate a frequency list, which was then used to select one hundred most frequently appearing specialist terms to be included in our legal English dictionary of collocations.

When it comes to establishing collocations, this type of software needs to use some statistical association measures that indicate the strength of association between two words. Such measures are based on various aspects of the co-occurrence of the two words in question and may be used to identify good collocation candidates, and it is basically up to the software developers as to which measures it offers its users. Until 2006, Sketch Engine used a MI-Score modified and AScore to give greater weight to the frequency of collocations. However, given the findings of more recent research, since 2006, it changed the statistic to logDice based on the Dice coefficient. In 2015, another adjustment was then made to the statistical methods applied in the software; the indices were modified again to compute the score more correctly. Currently, Sketch Engine applies a measure termed *logDice general word sketch score.* The score is computed for all possible word pairs, and the word pairs with the highest score are presented as collocation candidates.[5]

Having selected the terms, we analysed the collocational behaviour of each term individually using Word Sketch and its association measures. We have decided that each dictionary entry will include six types of collocations, i.e.

1) premodifier + TERM,
2) TERM + noun,
3) verb + TERM,
4) TERM + verb,
5) preposition + TERM,

---

[5] For detailed discussion, see https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/ and Rychlý (2008).

6)   TERM + preposition,

and each section in every entry is arranged in this manner; whenever no significant collocations are found in a given section, then the section is omitted. Moreover, each section starts with an alphabetical list of words which form a given type of collocation with the term. Another component of an entry is a sample sentence extracted from the corpus illustrating the use of a given collocation type, and the collocation itself is in bold so that it can be easily found by the dictionary user.

 A sample single section for the term 'court' is presented below:

premodifier + COURT
administrative, Appeal, appellate, appropriate, circuit, civil, commercial, County, Crown, divisional, domestic, European, federal, foreign, High, international, Lower, municipal, national, open, referring, sentencing, Supreme
The New York Times reports that "details in the article contradicted physical evidence already presented in **open court**."

It is worth noting that some words may collocate with several terms, and consequently, they will appear in the dictionary more than once. The dictionary will present the most frequent specialist terms together with their collocates in alphabetical order. We decided to set a minimum frequency threshold of five occurrences for a given collocate to be included in the main entry.

When it comes to some basic statistical data concerning the corpus created for this project, the corpus contains 9,505,800 tokens (the smallest units the corpus divides into), 8,098,719 words and 260,419 sentences culled from 636 documents representing UK Supreme Court Judgements. We hope that a corpus of such a size, created with high-quality documents from a reputable institution, will enable a reliable analysis of collocations appearing in legal English.

4. Conclusions

Despite some voices being raised as to the rationale behind a further increase in specialist dictionaries, there still seems to be a place for dictionary-like publications (although not necessarily in the form of a traditional book), as they meet a perennial human need, which is to learn a language so as to understand the world (L'Homme and Cormier 2014: 8). Thus, compiling a dictionary which will offer some insight into the concepts around which the legal system is organised may help the reader grasp the meaning of legal rules. In this article, we discussed the significance of collocations for understanding the intricacies of a specialist language and the assumptions and tools for preparation of a legal English dictionary of collocations. We believe that it is appropriate to perform further studies and that such a dictionary may serve as an aid for mastering the legal English genre. The final results, i.e. a dictionary of legal English collocations, will be presented and discussed in a separate article.

References

Bajčić, M. 2017. *New insights into semantics of legal concepts and the legal dictionary*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Biel, Ł. 2014. *Lost in the Eurofog: The textual fit of translated law*. Frankfurt: Peter Lang.

Evert, S. 2004. *Association measures*. www.collocations.de/AM/ [17/09/2019].

Evert, S. 2005. *The statistics of word cooccurrences: Word pairs and collocations*. Institutfür Maschinelle Sprachverarbeitung, Universität Stuttgart.

Evert, S. 2009. Corpora and collocations. In A. Lüdeling, and M. Kytö (eds.), *Corpus linguistics: An international handbook*, Vol. 2, 1212–1248. Berlin, New York: Mouton de Gruyter.

Fillmore, Ch. 1976. "Pragmatics and the description of discourse". In S. J. Schmidt (ed.), *Pragmatik/Pragmatics II. Grundlegungeinerexpliziten Pragmatic*, 83–104. Munich: Fink.

Gonzalez-Ray, I. 2002. *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.

Hausmann, F. J., and H. E. Wiegand. 1989. Components part and structures of general monolingual dictionaries: A survey. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, and L. Zgusta (eds.) *Wörtebücher. Ein internationals Hanbuch zur Lexikographie*, vol. 1, 328–360. Berlin/New York: de Gruyter.

Hausmann, F. J. 1997. Tout est idiomatique dans les langues. In M. Martins-Baltar (ed.), *La locution entre langue et usage*s, 227–290. Fontanay/Saint-Claud: ENS Éditions.

Hoey, M. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.

Kilgarriff A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel. 2014. The sketch engine: Ten years on. *Lexicography*, 1: 7–36.

Kjellmer, G. 1994. *A dictionary of English collocations*. Oxford: Clarendon Press.

Lehecka, T. 2015. Collocation and colligation. In J.-O. Östman, and J. Verschueren (eds.), *Handbook of pragmatics*, 1–20. John Benjamins Publishing.

L'Homme, M., and M. C. Cormier. 2014. Dictionaries and the digital revolution: A focus on users and lexical databases. *International Journal of Lexicography* 2014/ doi:10.1093/ijl/ecu023.

Mel'čuk, I. 1998. Collocations and lexical functions. In E. P. Cowie (ed.), *Phraseology, theory, analysis and application*, 23–53. Oxford: Clarendon Press.

Nielsen, S. 1994. *The bilingual LSP dictionary. Principles and practice for legal language*. Tübingen: Gunter Narr Verlag.

Patiño, P. 2014. Towards a definition of specialized collocation. In G. Quiroz Herrera, and P. Patiño Garcia (eds.), *LSP in Colombia. Advances and challenges*, 119–133. Brno: Peter Lang Publishing.

Rychlý, P. 2008. A lexicographer-friendly association score. *Proceedings of 2nd Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN*, 2: 6–9.

Šarčević, S. 1997. *New approach to legal translation*. The Hague/London/Boston: Kluwer Law International.

Siepmann, D. 2005. Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. *International Journal of Lexicography* 18(4): 409–443.

Siepmann, D. 2006. Collocation, colligation and encoding dictionaries. Part II: Lexicographic aspects. *International Journal of Lexicography* 19(1): 1–39.

Sinclair, J. 1966. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.), *In memory of J. R. Firth*, 410–430. London: Longman.

Sinclair, J. 2004. *Trust the text. Language, corpus and discourse*. New York/London: Routledge.

Wagner, A., and J.-C. Gémar. 2013. Materializing notions, concepts and language into another linguistic framework. *International Journal for the Semiotics of Law* 26: 731–745.

Wiechmann, D. 2008. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2): 253–290.

Online sources

supremecourt.uk

www.sketchengine.eu

Characterising the alternative and polar questions of Irish

Brian Nolan

Technological University Dublin, Ireland

Abstract

This paper gives an account of the similarities and differences between alternative and polar questions, where these question forms stand at the intersection of syntax, semantics, and pragmatics. We contrastively examine the nature of alternative and polar yes-no questions. We characterise the forms of these question types and the functions they serve. We examine the semantic and pragmatic dimensions of each question form and their answers. We characterise the felicity conditions necessary for their successful realisation of the speech act of requesting information via the alternative and yes-no interrogatives and assume that information is freely exchanged under a Gricean presumption of cooperation. We show that alternative questions have some similarities, but also significant differences, to polar yes-no questions. Alternative questions do not allow for yes-no answers. Instead, an appropriate answer must contain one of a selection from the alternative choice options listed in the framing of the question. Alternative questions are dependent on the presence of disjunction. We characterise the syntax and semantics of polar yes-no questions. We demonstrate in respect of the answers to polar yes-no questions of Irish that they contain instances of ellipsis and are full clausal expressions with a complete semantics where the elided elements are from the question part of the question-answer pair. The propositional content of polar yes-no questions is inferred from the context, specifically from the question with which the answer is paired. Irish does not have any exact words which directly correspond to English 'yes' or 'no' and so employs different strategies where a yes-no answer is required.

1. Introduction

This paper[1] compares and contrasts the syntactic forms of alternative questions with those of polar questions and how these question types differ in function, as instances of the speech act of requesting information. We characterise the syntax, functions, and pragmatics of alternate questions and polar yes-no interrogatives as question-answer pairs. We consider the felicity conditions necessary for the speech act interaction to be successfully realised. We assume that information is freely exchanged under a Gricean presumption of cooperation (Grice 1957, 1969, 1986). Irish uses a variety of syntactically significant question particles in the formation of its different question forms. From Dryer (2005: 470), we know that question particles are common

---

[1]   I wish to thank the two reviewers for their useful and helpful comments. The paper has benefited from these points raised.

in the world's languages as a cue in the syntactic structure of interrogatives. We examine the features of these in their morphosyntactic context. Other elements of the grammar of Irish have been reported on in Nolan (2008, 2012, 2013, 2014, 2017, 2019). We argue that a core function of interrogatives is the maintenance of common ground between the interlocutors. The construction and maintenance of common ground is crucial to retrieving meaning.

It is, of course, natural and normal during our discourses to seek information from our conversational partner. The canonical function of the interrogative is to seek information. Languages have different strategies for soliciting different kinds of information, and Irish is no different in this regard. Typically, these questioning strategies require sentences with different structural forms to act as signalling cues for the various kinds of information required (Nolan 2019:105–136). The use of questions and answers is therefore central within dialogue in language (Coulthard 1992, Dayal 2018, Dryer 2005, Hamblin 1973, Holmberg 2015).

Specifically, then, this paper gives an account of the differences between alternative and polar questions, where these question forms stand at the intersection of syntax, semantics, and pragmatics. The primary function of interrogatives is the maintenance of information in common ground via the update and exchange of information between the interlocutors. In this paper, we contrastively examine the nature of alternative and polar yes-no questions. We characterise the forms of these question types and the functions they serve, touching upon semantic and pragmatic dimensions of each question form and their answers. We characterise the felicity conditions necessary for the successful realisation of the speech act of requesting information via the alternative and yes-no interrogatives, and assume that information is freely exchanged under a Gricean presumption of cooperation.

We show that alternative questions have some similarities, but also significant differences, to polar yes-no questions. Alternative questions do not allow for yes-no answers. Instead, an appropriate answer must contain one of a selection from the alternative choice options listed in the framing of the question. Alternative questions are dependent on the presence of disjunction.

We demonstrate in respect of the answers to polar yes-no questions of Irish that they contain instances of ellipsis and are full clausal expressions with a complete semantics where the elided elements are from the question part of the question-answer pair. The propositional content is inferred from the context, specifically from the question with which the answer is paired. Irish does not have any exact words which directly correspond to English 'yes' or 'no' and so employs different strategies where a yes-no answer is required. The fact that languages have clausal types for asking questions shows clearly how important this activity is to human communication, and the construction and maintenance of shared common ground.

The paper has the following organisation. Section 2 examines the nature of the speech act of requesting information and outlines the felicity conditions necessary for a successful, i.e., felicitous, act of questioning. We note the important role of context in the meaning of these speech acts. In section 3 we examine Irish alternative questions while in Section 4 we explore the polar yes-no question form of Irish. Section 4 explores the nature of the answers to these question types, highlighting the main strategy used, the verb-echo strategy, when there are no direct words for 'yes' or 'no' in the language. We characterise the important features of the answers to the questions. Section 5 draws a number of conclusions and reviews similarities and differences between Irish alternative questions and polar questions.

2. Context and the speech act of requesting information

Interrogative sentences are associated with the speech act of requesting information via questioning. Interrogative sentences fall into three major classes depending on their syntactic and semantic properties. The required response to one of these question forms may be constrained in certain ways, depending on the nature of the question (1).

(1)   Types of questions
    a.   *Alternative questions* expect an answer based on one of the options presented within the question.
    b.   *Polar yes-no questions* expect affirmation or negation as an answer.
    c.   *Information questions (WH-questions)* typically expect an answer from an open set of possible replies, where the wh-word serves to target a specific information gap.[2]

In order to ask a question as a speech act, certain things must be appropriate in the context in which the speech act is uttered. In other words, a sentence must not only be grammatical to be correctly performed, it must also be felicitous. A central assumption is that the Gricean Cooperative Principle (Grice 1957, 1969, 1986) and its associated maxims apply. The Gricean Cooperative Principle Maxims are i) Quality: Speakers' contributions should be true; ii) Quantity: Speakers' contributions should be only as informative as the situation requires and speakers should refrain from saying either too little or too much; iii) Relevance: Contributions should relate to the purpose of the exchange; iv) Manner: Contributions should avoid obscurity and ambiguity and be clear, orderly and succinct. The speech act formalisation concerns the objective conditions of satisfaction for the speech act and its utterance meaning. In the discourse chain, the hearer's response to the question depends on whether the hearer believes that the speaker is obeying Grice's maxim of quality. If the hearer believes that the speaker *is* obeying the maxim of quality in asking the question, then the hearer can conclude that the speaker really does not know the answer. In this case, the most cooperative response is to simply inform the speaker of the answer, replying as appropriate to the particular question form. If the hearer believes that the speaker *is not* obeying the maxim of quality, then the hearer can conclude that the speaker does know the answer to the question, or that an answer is not actually required. In this case, the hearer can interpret the speaker's question *rhetorically*, as if it were meant only to bring the answer into the discourse. Actual felicity conditions apply to the question as a speech act, and these are described (2) where S = speaker; H = hearer; and p = some proposition.

(2)   Felicity conditions on the speech act of questioning
    FELICITY CONDITIONS: QUESTION
    PROPOSITIONAL CONTENT: Any proposition p
    PREPARATORY CONDITION: *Speaker* questions *Hearer* about proposition p *iff*
       (i)   S does <u>not know</u> the truth about p.
       (ii)   S <u>wants</u> to know the truth about p.
       (iii)   S <u>believes</u> H knows the truth about p that S wants.
       (iv)   It is not obvious that H will provide the information without being asked
    SINCERITY CONDITION: S wants this information
    ESSENTIAL CONDITION: Counts as an attempt to elicit this information

---

[2]   We do not examine information questions in this paper. See Nolan (forthcoming) for discussion.

A felicitous use of a question speech act requires that S: i) not be aware of the information requested and ii) S believes that H has knowledge of that information. As part of the felicity conditions of a question, we typically presuppose ignorance, or an information gap, on the part of the speaker, and a presumption of knowledge on the part of the addressee. Of course, a speaker may already possess the relevant information but can still felicitously pose the question to an addressee (as in some examination context), if the addressee's knowledge is considered to be in doubt. An utterance is identified as an interrogative questioning form because it has a certain morphosyntactic form that provides *cues* (Dryer 2005: 470–473). Consequently, it has an illocutionary force that requires an answer, or some response, from the addressee.

The importance of context in assisting in the determination of speech act meaning in language has long been recognised (Malinowski 1944; Nerlich 1990; Nerlich and Clarke 1996). According to Monaghan (1979:1), an analysis of the speech act with its context of situation is core and will assist us to: "in principle account for language in its social situation, rather than as merely a collection of structural units to be analysed individually", while more recently, Stalnaker (1999) has argued for the centrality of context in speech act meaning determination, and in support of the maintenance of common ground (Kecskes and Zhang 2009). Nolan (2014) characterised the cognitive states for an actor in the dialogue (3), while Nolan (2019) proposed a model of a situation that applied to speech acts along with several important aspects of speech act characterisation (4).

(3)   Cognitive states for an actor in a dialogue (Nolan 2014)
   a.   **bel**ieve' (Actor, p), has the meaning that the Actor believes that p is true for the Actor, where P is an expression in a human natural language.
   b.   **know'** (Actor, p) expresses a state of knowledge of the Actor with respect to p
   c.   **want'** (Actor, p) means that the Actor desires the event or state coded by p to occur.
   d.   **intend'** (Actor, p) means that the Actor intends to do p.

(4)   Important aspects of speech act characterisation (Nolan 2019)
   a.   The set of *beliefs* that the Actor has at any given time;
   b.   The *goals* that the Actor will try to achieve;
   c.   The *actions* that the Actor performs and
   d.   The *knowledge of the effects* of these actions;
   e.   The *environment information / knowledge* that the Actor has (which may be incomplete or incorrect);

Context has a central role as a component of cognition in the determination of the conditions of knowledge activation as well as which elements of our knowledge apply in a given situation. Context is activated, and constructed, in the ongoing interaction as it becomes relevant, and is eventually shared by discourse interlocutors in the construction of the discourse common ground. Context includes cultural knowledge, general knowledge and shared communal beliefs, and the general experience that arises from the interplay of culture and social community. Context may also include location and environment. In determining the meaning of a speech act utterance, situational and contextual factors need to be assessed. Here we consider alternative questions and polar yes-no questions, as they are found in Irish.

3. The question particles

To pose a simple alternate or yes-no question, certain question particles are used that are common to both question forms (Table 1). The forms of the question particle can vary according to the morphosyntactic context and have morphosyntactic consequences (see Table 2). On occasion, in casual speech the question particle may be omitted.

More specifically, the copula is described by Doherty (1996:11) as an inflectional particle that has a paradigm for combination with various 'presentential particles' (5). Advancing our understanding of how copula sentences relate to verbal sentences, Ó Sé (1990:2–4) considers how various verbal grammatical categories are represented in the Irish copula, with only two basic forms of the copula occurring in Modern Irish, *is* and *ba*, that fuse with negative and interrogative particles. Linguists are not yet in agreement regarding the status of the Modern Irish copula and its character.

(5)	**Present/Future**	**Past/Conditional**
interrogative + copula	*an*	*ar / arbh*
negation + copula	*ni/chan*	*nior(bh) / char(bh)*
subordination + copula	*gur / gurb*	*gur / gurbh*

Table 1: The appropriate question particle

present tense	**affirmative:**	***An*** + **verb** + **subject** + **object₁** + **object₂** *An ólann tú fuisce nó caife?* Do you drink whiskey or coffee?	Alternative question
	negative	***Nach*** + **verb** + **subject** + **object** *Nach n-ólann tú fuisce?* Don't you drink whiskey?	Polar question
past/preterit tense	**affirmative:**	***Ar*** + **verb** + **subject** + **object₁** + **object₂** *Ar ól tú fuisce nó caife?* Did you drink whiskey or coffee?	Alternative question
	negative	***Nár*** + **verb** + **subject** + **object** *Nár ól tú fuisce?* Didn't you drink whiskey?	Polar question

Table 2: The morphosyntax of the question particle

Tense	Affirmative particle	Lenition/Eclipsis[3]	Negative particle	Lenition/Eclipsis
Present	*an*	E	*nach*	E
Past/preterit	*ar*	L	*nár*	L

---

[3] Lenition, (called *séimhiú* in Irish) is an initial mutation that affects the spelling and pronunciation of words that begin with the letters *b, c, d, f, g, m, p, s* and *t*. The word meaning does not change. Lenition is represented in written text with a letter *h* placed after the first letter of a word. Lenition is also referred to as aspiration. (Source: http://www.nualeargais.ie/gnag/lenition.htm).

Eclipsis, also known as nasalisation, is the substitution of the unvoiced consonants with the corresponding voiced ones (*c* becomes *g*, *f* becomes *bh*, *t* becomes *d*, *p* becomes b). Voiced consonants are then replaced with nasals (*b* becomes *m*, *d* becomes *n*, *g* becomes *ng*).

We are interested in the interrogative form. While the question particle *an* is used as an alternate or polar question marker[4], the particle *an* can be used to support a variety of simple queries (6).

(6)   Introducing a simple query

    a.   *Ar   chuir   tú   isteach   air?*
        QPST put.PST 2SG in   on+3SG
        Did you apply for it?

    b.   *An   ndéanfaidh   siad   é?*
        Q   do.FUT   3PL   3SG.M.ACC
        Will they do it?

    c.   *Ar   bhuail   sé   é?*
        QPST hit   3SG.M.NOM 3SG.M.ACC
        Did he hit him?

    d.   *An   dochtúir   é?*
        QCOP doctor   3SG.M.ACC
        Is he a doctor?

    e.   *An   leatsa   an carr sin?*
        QCOP with+2SG det car   that
        Do you own that car?

    f.   *An   fear mór a   bhí   ann?*
        QCOP man big REL be.PST there
        Was he a big man?

    g.   *Ar   fada go   mbeidh deiread leób?*
        QCOP long that be.FUT end   with+3PL
        Will it be long before there's an end to them?

4. The alternative questions of Irish

Alternative questions are similar in many ways to polar yes-no questions, but also have important differences as we will see. A major difference is that alternative questions do not, under any circumstances, allow for yes-no answers. Alternative questions are constructed in such a way that the licenced answer cannot be yes, or no. Instead, an appropriate answer must contain one of a selection from the alternative choice options listed in the framing of the question. Alternative questions are dependent on the presence of disjunction. Consequently, the set of alternative propositions framed within the question is equal to the set of possible

consonant	eclipsed	spoken	consonant	eclipsed	spoken
b	mb	[m]	g	ng	[ng]
c	gc	[g]	p	bp	[b]
d	nd	[n]	t	dt	[d]
f	bhf	[v]/[w]			

(Source: http://www.nualeargais.ie/gnag/eklipse.htm)

[4] Data is used from the websites:
https://en.wikipedia.org/wiki/Irish_conjugation#Interrogative_particles,   https://www.teanglann.ie/ga/,
https://www.gaois.ie/en/corpora/monolingual/, http://www.nualeargais.ie/gnag/kopul1.htm.
Scholarly works consulted include: *An caighdeán oifigiúil* 2017; Bennett, *et al.* 2015; Christian Brothers 1997; de Bhaldraithe 1987; Collins Irish Grammar 2011. Doherty 1996; Doyle 2001; McCloskey 1991; McGonagle 1991; Ó Dónaill 1981; Ó Mianáin 2020, Ó Sé 1990; Ó Siadhail 1989, Ward 1974.

answers defined in virtue of the form of the question. A primary quality of alternative questions then is that they typically only allow felicitous answers which choose one out of the set of alternatives listed in the question (or by declaring a lack of interest to any of the listed alternatives). That is, alternative questions list the set of alternatives as option-1, option-2, and so on, within the formulation of the question. Specifically, in the alternative question, the allowed alternatives are explicitly stated within the question clause. These alternative questions of Irish typically take the schematic format in (7).

(7)　　Alternative question format

an	*...*	NP$_X$	*nó*	NP$_Y$
QPRT	…	X	or	Y

The alternative question (8) posed by the speaker is used to determine which of the two (or possibly more) specific alternatives holds for the addressee. The alternative question may be posed in different ways but the format of the answer remains the same. In alternative questions, a set of alternatives is proposed by the question and a selection from these is requested from the interlocutor (9) by way of a response. Before the question is answered, S is unclear as to whether H wants option-1 or option-2. For convenience, we label these options as Y1 and Y2. Only one of these is selected by H within the answer. The logical structure and internal semantics of the alternative question and possible answers is represented in (10–11). The elided elements are flagged in the logical structure representation of the asserted answer – these are in common ground as referents because they were introduced within the question.

Alternative question

(8)　a.　*An　fearr　leat　　　tae　nó　caife?*
　　　　　QPRT　prefer　with:prep+2SG　tea　or　coffee
　　　　　Do you prefer tea or coffee?

　　b.　*An　ólann　　tú　tae　nó　caife?*
　　　　　QPRT　drink:PRS　2SG　tea　or　coffee
　　　　　Do you drink tea or coffee?

Alternative question

(9)　a.　Q:　　　　*An　ólann　　tú　tae　nó　caife?*
　　　　　　　　　　QPRT　drink:PRS　2SG　tea　or　coffee
　　　　　　　　　　Do you drink tea or coffee?
　　　　　　　　　　P=you drink tea or coffee

　　　　Answer 1:　*Ólaim　　　tae.*　　　　　　　　　[Positive response]
　　　　　　　　　　drink:V.PRS　　tea
　　　　　　　　　　I drink tea.

　　　　Answer 2:　*Tae.*　　　　　　　　　　　　　　[Positive response]
　　　　　　　　　　Tea.
　　　　　　　　　　([I drink]$^{\text{ellipsis}}$) tea.

　　b.　Q:　　　　*An　ólann　　sí　tae　nó　caife?*
　　　　　　　　　　QPRT　drink:PRS　3SG.F　tea　or　coffee
　　　　　　　　　　Does she drink tea or coffee?
　　　　　　　　　　P=she drinks tea or coffee

　　　　Answer 1:　*Ólann　　sí　tae.*　　　　　　　　[Positive response]
　　　　　　　　　　drink:V.PRS　3SG.F　tea
　　　　　　　　　　She drinks tea.

Answer 2:   *Tae.*   [Positive response]
              Tea.
              ([she drinks]^ellipsis) tea.

As a response to the question, the answer is a speech act of assertion. The logical structure of the alternative question and answer is indicated in (10); the X argument is not spoken, it is elided, but is retrievable from context. The semantics of the question is indicated in (11). Neither the verbal predicate nor the X argument is spoken, both are elided, but again retrievable from context. With alternative questions, the disjunctive set of possible responses is constrained by the list of alternative options that are specifically listed in the question. The hearer picks one of these options in the construction of the answer.

(10)   Logical structure of the alternative question and answer
       a.   **Q.**$_{Alternative}$ [**do'** (x, **pred'**(x, y1 ∨ y2))]
       b.   [**assert'** (**do'** (x, **pred'**(x, y1)))]
       c.   [**assert'** (**do'** (x^{ellipsis}, **pred**^{ellipsis}**'**(x^{ellipsis}, y1)))]

(11)   Semantics: Question' (P=you drink tea or coffee, tea | coffee)
       context before question
       S: **know'** (p, ?)
       context after answer
       S: **Bel'** (p, tea)   or S: **Bel'** (p, coffee)
       S: **know'** (p, tea) or S: **know'** (p, coffee)

Subject to pragmatic context, there is a reading of an alternative question that licences a polar question interpretation (12). In this context, the [*tae nó caife*]$_{np}$ is viewed as a single complex NP rather than a list of choices. In addition, it is possible to view the alternative question as a disjunction of two (or more) polar questions (13). In this view, in the second disjunct, the question particle, verbal predicate and subject argument are all elided in the syntactic realisation.

(12)   a.   Alternative Q: *An   ólann   tú   [tae]$_{np}$ nó [caife]$_{np}$?*
       b.   Polar Q: *An   ólann   tú   [tae   nó   caife]$_{np}$?*
                               QPRT   drink:PRS 2sg   [tea   or   coffee]
                               Do you drink [tea or coffee]?
                               P=you drink [tea or coffee] | ¬p

(13)   Disjunct Polar Q: *An   ólann   tú   tae   nó   [an   ólann   tú]* $_{elided}$ *caife?*
                               QPRT   drink:PRS 2SG   [tea]   or   [QPRT drink:prs 2SG] $_{elided}$ coffee
                               Do you drink tea or [do you drink] $_{elided}$ coffee?
                               P=you drink tea | you drink coffee | ¬p

We therefore have two perspectives on the nature of the alternative question form. In the first view, we can treat the alternative question as a distinct interrogative form containing a list of options (two or more) from which the addressee, in their answer, must make a single selection. In the second view, we can consider the alternative question as a realisation of the polar question but reflecting disjunction across two (or more) clauses within the sentence. Ellipsis occurs in

the second clause. Here, the negative proposition is an implicit 'option'. This second view considers the alternative question form as underlyingly disjoined polar questions.

With alternative questions:

a. A specific question particle is used.
b. A specific syntactic form is employed with alternative questions.
c. The alternative question form contains a list of options such that the answer must select from one of these.
d. The alternative question form may involve instances of ellipsis.
e. The answer to an alternative question contains ellipsis of the material following the verb.
f. The answer to an alternative question may optionally contain ellipsis of the verbal predicate as well as the subject argument.
g. The alternative question form can be viewed as a realisation of two (or more) disjunct polar questions.

We now proceed, in the next section, to examine polar yes-no questions of Irish.

5. The polar yes-no question as an interrogative sentence

We have considered alternative questions as an interrogative form where a set of alternatives is proposed and a selection from amongst those is requested from the interlocutor. In contrast, polar yes-no questions denote a set consisting of a proposition and its negation and request confirmation or negation of the proposition. Polar yes-no questions are typically used to inquire about the truth or falsity of the proposition they express. As well as a positive confirmation or negative type answer, answers to polar questions can, of course, typically also assume any value on a scale between 'true' and 'false', as, for example, 'maybe', 'I don't know', 'perhaps', 'possibly', and other such responses.

Essentially, a polar question has a set of two propositions, p and ¬ p, but frames just one of these, p, in the question. The ¬ p option is implicit and salient but not syntactically stated. A polar question cannot denote a singleton set and the negative proposition is always (implicitly) available. There are two syntactic constructional forms (14) and (15) for the polar yes-no question, the first form has a matrix lexical verb while the second being a copula form. Across both forms of polar yes-no questions, the morphosyntax requires the clausal form to have the question particle at the front of the clause.

[Polar question – lexical verb form]
(14) Q: *An   ólann   sé   tae?*
         QPRT   drink:V.PRS 3SG.M tea
         Does he drink tea?
         Proposition: [p = he drinks tea]
         Negative of Proposition: [¬p = he does not drink tea]

[Polar question – copula form]

(15) Q: *An   dochtúir   í   Aifric?*
           QCOP   doctor   3SG.F.ACC Aifric
           Is Aifric a doctor?
           Proposition: [p = Aifric is a doctor]
           Negative of proposition: [¬p = Aifric is not a doctor]

A feature of the polar yes-no question is that it can be *biased* towards a particular polarity, positive or a negative, within the answer. This happens through the use of forms such as 'someone', for example, within the question, which bias the question towards a positive orientation, as in (16).

(16) Q: *Ar   thug   éinne   cuairt   aréir?*
           QPST   give.V.PST anyone   visit   last night
           Lit: Did anyone give a visit last night?
           Did anyone visit last night?
           Proposition: [p = someone visited last night]

This indicates that the speaker believes that the answer is positive and is seeking confirmation. A polar yes-no question specifically favours the core proposition framed by the question. This is the property of polar questions that makes them amenable to a yes-no response. Polar questions therefore make salient the affirmative, along with the implicit negative version of a proposition. The truth-conditional aspect of the meaning of a polar question is simply the set of the truth-conditional meanings of the possible answers to the question. Pragmatically, however, a polar question may also reveal information about the speaker's bias towards a particular answer. The speakers' bias may be with regard to evidence present in the conversational context, that is, an evidential bias. Any contextual evidence is mutually available to the participants in the discourse situation and typically forms part of the present shared common ground. Evidential bias is about contextual information available to all conversational participants, in the shared common ground and inherently public. In contrast, any epistemic bias that may occur is grounded on the speaker's own private beliefs, and need not be shared by other conversational participants.

Typically, we expect that some contextual evidence influences what the speaker believes. Also, posing a question requires the hearer to identify the questioner's intention. Given the context dependency, the evidential or epistemic bias of a polar question assists the hearer in determining what type of information the questioner is seeking. Specifically, the polar question establishes a relation between the propositional content of the question and the speaker's attitude, and whether the proposition p is in the speaker's belief set.

The strategy that Irish employs in answers to the first form of the polar yes-no questions (with a lexical verb) is a verb-echo strategy. In many languages (Dryer 2005), yes-no questions are typically answered, not by an affirmative or negative particle, but by echoing the matrix verb of the question for positive answers and echoing the verb of the question plus a negation marker of some kind for negative answers. An affirmative answer is an echo of the matrix lexical verb in the question, while the negative answer is an echo of the matrix lexical verb in the question combined with a sentential negation.

We can consider the structure of these answers in more detail. Within the answer in Irish to the first form of the polar yes-no questions, the tense-marked verb form is used with an explicit nominal argument (17). When a synthetic verb form is used, a pronominal appears in the grammatical relation of nominative subject within the answer. In the synthetic form, the PN is conflated on the end of the verb as a suffix. Additionally, in negative polarity answers, the negative particle is also used.

(17) Q: *An*   *ólann*   *tú*   *bainne?*
          QPRT drink:V.PRS 2SG.M milk
          Do you drink milk?
      Answer 1: *Ólaim.*   [Positive response]
                 Drink:V.PRS+1SG
                 Lit.: drink I
                 Yes.
      Answer 2: *Ní*   *ólaim.*   [Negative response]
                 NEG   drink:V.PRS+1SG
                 Lit.: not drink I
                 No.

A grammatical subject may be used when the speaker chooses an emphatic affirmation or denial (18 Answer 2, in comparison to 18 Answer 1).

(18) Q: *An*   *ólann*   *sé*   *uisce?*
          QPRT drink:V.PRS 3SG.M water
          Does he drink water?
      Answer 1: *Ní*   *ólann*   [Negative response]
                 NEG   drink:V.PRS
                 Lit.: Not drink
                 No.
      Answer 2: *Ní*   *ólann*   *sé*   <u>*ar chor ar bith!*</u>   [Negative emphatic response]
                 NEG   drink:V.PRS 3SG.M at all
                 Lit: He doesn't drink at all!
                 No

An answer to the first form of the polar yes-no questions can allow for a non-specific response indicating a lack of precise knowledge (19).

(19) Q: *An*   *raibh*   *na*   *feirmeoirí* *orgánacha* *sásta?*
          QPRT be.AUX.PST DET.PL farmers   organic   satisfied
          Were the organic farmers satisfied?
      Answer 1: *Bhí*   *siad.*
                 Be.AUX.PST 3PL
                 They were (satisfied)!
      Answer 2: *Níl*   *a*   *fhios*   *agam.*
                 NEG   REL   know   at:PREP+1SG
                 Lit: knowledge is not at me
                 I don't know.

Additionally, a response indicting uncertainty allows elements of the actual question to be embedded in the response (20).

(20)   Q: *An   mbeidh   mórán daoine   anseo?*
            QPRT be.AUX.FUT many   people   DET+here
            Will there be many people here?
       Answer: *Níl   mé   cinnte   an   mbeidh*
                 NEG   1SG   certain QPRT   be.AUX.FUT
                 mórán   daoine   anseo.
                 many   people   DET+here
                 I am not certain how many people will here.

An answer to the first form of a polar yes-no question has several characteristics worthy of note:

   a.   The lexical verb within the answer is echoed from the question and inflected for tense. As tense is a clausal category that locates the time of the event denoted by a clause in relation to the time of utterance, this is evidence that the answer is a clause.
   b.   When it occurs in the answer to a polar yes-no question, pronominal subject marking implies the presence of a subject, hence also the presence of a clause.
   c.   In the answer to a polar yes-no question, ellipsis occurs with respect to the arguments of the verb. Ellipsis can also occur with respect to the verb itself in an alternative question form, as we have seen with example (9).
   d.   The VSO word order of the answer is maintained and a pronominal constituent is case-marked for the grammatical function of subject in the canonical position post-verb.
   e.   The context to a polar yes-no question, before the question is answered, has the speaker S uncertain as to whether the proposition is true, or false. Only one of $(p \mid \neg p)$ holds.

The evidence supports the view that answers to polar yes-no questions have a sentential structure even when they consist of just one pronounced word. The elided material is in common ground and retrieved when needed by the hearer H for meaning resolution. Specifically, the elided (unpronounced) material is retrieved from the clausal content in the question. A representative example of the logical structure of the first form of the polar question (21) and a typical answer is indicated in (22), with associated semantics is indicated in (23).

       Polar question
(21)   Q: *An   ólann   sé   bainne?*
                    QPRT   drink:V.PRS 3SG.M   milk
                    Does he drink milk?
                    p=you drink milk | you do not drink milk
       Answer 1:   *Ólann.*   [Positive response]
                    drink:V.PRS
                    Lit.: drink
                    Yes.

(22)   Logical structure of the polar question and answer
       a.   **Q.$_{\text{polar}}$ [do'** (x, **pred'**(x, y)]
       b.   [**assert'** (**do'** (x^{ellipsis}, **pred'**(x^{ellipsis}, y^{ellipsis})))]

(23)   Semantics: Question' (P=he drinks milk, p | ¬p)
                     context before question
                     S: **Bel'** (p, *true | false*)
                     S: **know'** (p, ?)

context after answer

S: **Bel'**(p, *true* | *false*)

S: **know'**(p, *true* | *false*)

We now briefly look at the second form of the polar question, the copular form (24). An interesting fact of Irish is that the language does not have any exact words which directly correspond to English *yes* or *no* and so the language necessarily employs a different strategy where a yes-no answer is required. To formulate an answer to the copula-form polar question, equivalent to a yes or no of English, the copula-derived phrases *sea* (cop+3sg = 'be-it') and *ní hea* (neg.cop *3sg* = 'neg be it') are used. These function as logically equivalent to 'yes' and 'no', respectively. The *sea / ní hea* can be used in formulating a response to a question with either a M or F referent (24a vs. 24b). Like the first form of the polar question with the lexical verb, this second form also inquires after the truth value of a proposition.

In these question forms, the copula is used with an interrogative purpose, to elicit information from the partner interlocutor. With an interrogation function, the copula clause uses a different questioning structure for classification vs identification, based on the copula structure indicated in (25). Examples of the copula classification and identification interrogative functions are found in (26) – (27).

[Copula construction]

(24) a. Q: *An   dochtúir   é   Lorcán?*
                 QCOP   doctor   3SG.M.ACC Lorcan
                 Is Lorcan a doctor?

      b. Q: *An   dochtúir   í   Aisling?*
                 QCOP   doctor   3SG.F.ACC Aisling
                 Is Aisling a doctor?

          Answer 1:   *Sea.* (= is$_{cop}$+ea$_{3sg.neut}$)   [Positive response]
                       COP+3SG.NEUT
                       Lit: Be (s)he.
                       Yes.

          Answer 2:   *Ní   hea.*   [Negative response]
                       NEGCOP   3SG.NEUT
                       Lit: Not be (s)he.
                       No.

(25) The general structure of the Irish Copula clause
      a)   Classification: cop <u>Predicate</u> Subject
      b)   Classification: cop <u>Predicate</u>$_{part-1}$ Subject <u>Predicate</u>$_{part-2}$
      c)   Identification: cop <u>Predicate</u> Subject$_1$
      d)   Ownership Identity: cop [<u>Preposition *le* 'with' + NP</u>]$_{Predicate}$ Subject
      e)   Emphasis: cop <u>Predicate</u> Subject$_1$
           Where
                <u>Predicate</u>$_{part-2}$ may contain a relative clause
                Subject$_1$ may be either [NP] or [NP + a relative clause]

Copula sentences are essentially equational units which establish an identity between a known or presupposed piece of information or entity, and a focused entity that presents new information of some kind. The copula of Modern Irish, as a marked focusing construction that

brings particular terms into focus, is examined in Nolan (2012: 192–217). These terms may be an NP, an adverb, an adjective, an adposition or other constituent including the verb. The copula predicate contains the new focused information and the copula subject contains the given topic. Whenever a speaker delivers an utterance in a particular context it is done with communicative intent, and the addressee must be able to unpack the utterance given the shared context.

A speaker has to distinguish three facets of knowledge with respect to the hearer of the utterance: a) That which is computable from the physical context; b) That which is available from what has already been said; c) That which is available from common ground knowledge. In an identifying copula, which takes the generalised schematic form:[x is y], the identifying expression in focus is the 'x', and this is most likely to be definite ([def +]). The 'y' term is the topic of the construction and represents the entity or information already available to the addressee. With the alternative and polar question forms, the copula is used with an interrogative purpose.

Typically, the classification copula sentence (with copula + indefinite noun + pronoun) will require an indefinite noun while the copula in the identification function (with copula + pronoun/definite noun + pronoun/definite noun) will have a subject that is both specific and definite (Ó Siadhail 1991:224–225). The Collins Irish Grammar (2011:114–115) informs us that *an* changes to *ar* for past tense and lenites/aspirates the predicate, but does not change before vowels. In turn, *ar* correspondingly changes to *arbh* for past tense before the pronouns *é*, *í* and *iad*. The negative forms apply here also (identification: *nach-> nár*, classification: *nach-> nárbh*).

(26) a. *An   duine   cliste   é?*   **Classification function**
          Q<small>COP</small>   person clever   3<small>SG.M.ACC</small>
          Is he a clever person ?

      b. *An   amharclann í?*
          Q<small>COP</small>   theatre   3<small>SG.F.ACC</small>
          Is it a theatre?

      c. *Ar   dhuine   cliste   é?*
          Q<small>COP.PST</small>   person clever   3<small>SG.M.ACC</small>
          Was he a clever person ?

      d. *Ar   amharclann í?*
          Q<small>COP.PST</small> theatre   3<small>SG.F.ACC</small>
          Was it a theatre?

(27) a. *An   é   Dónal   do   chara ?*   **Identification function**
          Q<small>COP</small>   3<small>SG.M.ACC</small> Dónal   your   friend
          Is Dónal your friend?

      b. *An   í   Aisling   an   cailín fionn?*
          Q<small>COP</small>   3<small>SG.F.ACC</small>   Aisling   DET   girl   blond
          Is Aisling the blonde girl?

      c. *Arbh   é   Dónal   do   chara?*
          Q<small>COP.PST</small>   3<small>SG.M.ACC</small> Dónal   your   friend
          Was Dónal your friend?

We now briefly consider the rhetorical polar yes-no question. The function of polar questions is to elicit confirmation. Some polar yes-no questions have a rhetorical force (28) that arises when conditions supporting the felicitous use of the speech act of questioning are not met. These questions do not seem to expect a positive or negative answer and are rhetorical in nature. Indeed, responses to the rhetorical polar questions would go against common ground knowledge.

(28) *Ní féidir leat mo cheist a fhreagairt, an féidir?*
      Ní   féidir leat   mo cheist   a fhreagairt, an   féidir?
      NEG able   with:prep+2SG my question   REL answer,   QPRT able
      You can't answer my question, can you?

Rhetorical polar yes-no questions are actually a particular kind of information seeking questions that masquerade as a polar form but they do not facilitate a yes or no answer. They contain cues from syntax and context which alerts us that the speaker is taking a rhetorical stance.

6. Concluding discussion

Irish has clausal types with a specific syntax for requesting information through alternative and polar yes-no questions, like other languages of the world (Haspelmath, Dryer, Gill, and Comrie 2005; Dryer 2005). Asking questions is a core human activity central to communication and understanding. The construction and maintenance of common ground through questioning facilitates the construction of meaning between the interlocutors.

Alternative questions do not inquire about the truth value of a proposition, but rather ask which listed option among a selection of alternatives is most appropriate in a given context. With an alternative question, knowing the meaning of the question is knowing its possible answer. Polar questions do not overtly provide a selection of answer possibilities in either of the forms found in Irish.

With alternative questions we found that:

a. A specific question particle is used.
b. A specific syntactic form is employed with alternative questions.
c. The alternative question form contains a list of options such that the answer must select from one of these.
d. The alternative question form may involve instances of ellipsis.
e. The answer to an alternative question contains ellipsis of the material following the verb.
f. The answer to an alternative question may optionally contain ellipsis of the verbal predicate as well as the subject argument.
g. The alternative question form can be viewed as a realisation of two (or more) disjunct polar questions.

The polar yes-no question is a disjunction of p and ¬p. The purpose of the polar question is about confirming the truth value of a proposition. It is also about managing the belief status

of knowledge in common ground. It also updates the information gap in the speaker's common ground from information provided within the answer. Common ground acts as a kind of decentralised knowledge system supporting the cognitive activation of relevant contextual knowledge. Context contributes to the meaning resolution, and to the resolution of any evidential or epistemic bias that may exist.

The answers to polar yes-no questions of Irish contain instances of ellipsis and, as such, represent full clausal expressions with a complete semantics where the elided elements are from the question part of the question-answer pair. To formulate an answer to the first form of the polar yes-no question, the grammar of Irish uses a verb-echo strategy whereby the matrix verb of the polar yes-no question is used in the answer. Nominal argument ellipsis of one or two arguments also occurs. The success of this strategy requires the construction and maintenance of common ground for the retrieval of elided arguments.

We found that the answer to the first form of a polar yes-no question has several significant characteristics:

a. The lexical verb within the answer is echoed from the question and inflected for tense. As tense is a clausal category that locates the time of the event denoted by a clause in relation to the time of utterance, this is evidence that the answer is a clause.

b. When it occurs in the answer to a polar yes-no question, pronominal subject marking implies the presence of a subject, hence also the presence of a clause.

c. In the answer to a polar yes-no question, ellipsis occurs with respect to the arguments of the verb, but not the verb itself.

d. The VSO word order of the answer is maintained and a pronominal constituent is case-marked for the grammatical function of subject in the canonical position post-verb.

e. The context to a polar yes-no question, before the question is answered, has the speaker S uncertain as to whether the proposition is true, or false. Only one of (p | ¬p) holds.

The evidence is that the propositional content in a question-answer interaction is inferred from the context, specifically from the question with which the answer is paired.

References

An caighdeán oifigiúil. 2017. *Gramadach na Gaeilge: An caighdeán oifigiúil, an dara cló.* Dublin: Seirbhís Thithe an Oireachtais. Available at https://data.oireachtas.ie/ie/oireachtas/caighdeanOifigiul/2017/2017-08-03_an-caighdean-oifigiuil-2017_en.pdf. (29 Sept 2020)

Bennett, R., Elfner, E., and J. McCloskey. 2015. *Prosody, focus and ellipsis in Irish.* MS. Available at: https://www.linguisticsociety.org/sites/default/files/04_95.1Bennett.pdf. (29 Sept2020)

Christian Brothers, The. 1997. *New Irish grammar.* Dublin: C.J. Fallon, Mount Salus Press.

Comer, N. 2011. *Collins Irish grammar.* Glasgow, UK: Harper Collins Publishers.

Coulthard, M. (ed.). 1992. *Advances in spoken discourse analysis.* London: Routledge.

Dayal, V. 2018. *Questions.* Cambridge: Cambridge University Press.

de Bhaldraithe, T. 1987 [1959]. *English-Irish dictionary (with terminological additions and corrections).* Dublin: An Gúm.

Doherty, C. 1996. Clausal structure and the Modern Irish copula. *Natural Language & Linguistic Theory* 14(1): 1–46.

Doyle, A. 2001. *Irish.* Languages of the World/Materials Series. Muenchen: LINCOM Europa.

Dryer, M. S. 2005. Polar questions. In M. Haspelmath., M. S. Dryer, D. Gill, and B. Comrie (eds.), *The world atlas of language structures*, 470–473.Oxford: Oxford University Press.

Grice, H. P. 1957. Meaning. *Philosophical Review* 66, 377–388. Reprinted in P. F. Strawson (ed.). *Philosophical logic*. London, 1967.

Grice, H. P. 1969. Utterer's meaning and intentions. *Philosophical Review* 78: 147–177. Reprinted in Grice, H. P. 1975. Logic and conversation. *Syntax and Semantics 3: Speech Acts*, ed. by Peter Cole and Jerry L. Morgan. 41–58. New York: Academic Press.

Grice, H. P. 1986. *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Hamblin, C. L. 1973. Questions in Montague English. *Language* 10(1): 41–53.

Holmberg, A. 2015. *The syntax of yes and no*. Oxford: Oxford University Press.

Kecskes, I. and F. Zhang. 2009. Activating, seeking, and creating common ground: A socio-cognitive approach. *Pragmatics & Cognition* 17(2): 331–355.

Malinowski, B. 1944. *A scientific theory of culture and other essays*. Chapel Hill: The University of North Carolina Press.

McCloskey, J. 1991. Clause structure, ellipsis and proper government in Irish. *Lingua* 85: 259–302.

McGonagle, N. 1991. *Irish grammar: A basic handbook*. Indreabhan, Conamara, Galway: Cló Iar-Chonnachta.

Monaghan, J. 1979. *The Neo-Firthian tradition and its contribution to general linguistics*. Tubingen: Max Niemeyer Verlag.

Nerlich, B. 1990. *Change in language: Whitney, Bréal, and Wegener*. London and New York: Routledge.

Nerlich, B and D. D. Clarke. 1996. *Language, action, and context – The early history of pragmatics in Europe and America, 1780–1930*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nolan, B. 2008. Modality in RRG: Towards a characterisation using Irish data. In R. Van Valin, (ed.), *Investigations of the syntax-semantics-pragmatics interface. Studies in language companion series 105*, 147–159. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nolan, B. 2012. *The syntax of Irish: A functional account*. Sheffield: Equinox.

Nolan, B. 2013. Constructions as grammatical objects: A case study of the prepositional ditransitive construction in Modern Irish. In B. Nolan and E. Diedrichsen (eds.), *Linking Constructions into functional linguistics. Studies in language companion series 145*, 143–178.Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nolan, B. 2014. Extending a lexicalist functional grammar through speech acts, constructions and conversational software agents. In B. Nolan and C. Periñán (eds.), *Language processing and grammars: The role of functionally oriented computational models. Studies in language companion series 150*, 143–164. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nolan, B. 2017. The syntactic realisation of complex events and complex predicates in situations of Irish. In B. Nolan and E. Diedrichsen (eds.), *Argument realisation in complex predicates and complex events. Studies in language companion series 180*, 13–41. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nolan, B. 2019. The role of context and common ground in utterance meaning with assertive and declarative speech acts of Irish. In A. Bondaruk and A. Bloch-Rozmej (eds.). *Studies in formal linguistics: Universal patterns and language specific parameters*. Berlin: Peter Lang.

Nolan, B. 2019. The forms, functions and pragmatics of Irish polar questions and answers. In B. Nolan, and E. Diedrichsen (eds.). *Linguistic perspectives on the construction of meaning and knowledge*, 105–136. Newcastle upon Tyne: Cambridge Scholars Publishing.

Nolan, B. [forthcoming]. The syntax, semantics and pragmatics of Irish Information questions. In K. Drabikowska and A. Prażmowska (eds.), *Exploring variation in linguistic patterns*, 175–196. Lublin: Wydawnictwo KUL.

Nolan, B., and E. Diedrichsen. 2013. *Linking Constructions into functional linguistics. Studies in language companion series 145*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ó Dónaill, N. 1981. *Gearrfhochlóir Gaeilge–Béarla*. Dublin: An Roinn Oideachais, Oifig an tSoláthair.

Ó Mianáin, P. 2020. *Concise English–Irish dictionary (An Foclóir Nua Béarla–Gaeilge)*. Baile Átha Cliath: An Gúm, Foras na Gaeilge.

Ó Sé, D. 1990. Tense and mood in Irish copula sentences. *Ériu* 41: 61–75.

Ó Siadhail, M. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge MA: Cambridge University Press.

Stalnaker, R. 1999. *Context and content*, Oxford: Oxford University Press.

Ward, A. 1974. *The grammatical structure of Munster Irish*. Ph.D. diss. Trinity College Dublin.

Fluency and complexity as coupled growers in speaking English at secondary school – A case study of a good, average, and poor language learner

Katarzyna Rokoszewska

Jan Długosz University in Częstochowa, Poland

Abstract

One of the main assumptions of Complex Dynamic Systems Theory (CDST) is that internally complex language subsystems develop non-linearly while entering different kinds of supportive, competitive, conditional, or dual relationships which are characterised by trade-offs caused by learners' restricted cognitive processing, especially in foreign language speech. The present paper belongs to a short series of articles which examines various aspects of the development of L2 English speech at secondary school on basis of the same longitudinal, exploratory, and corpus-based case study. The aim of this paper is to investigate the dynamics of the relationships between fluency and both syntactic and lexical complexity in the speech of a good, average, and poor language learner at the level of secondary school. Syntactic complexity was investigated in terms of general sentence complexity, subordination, coordination, and nominalisation, whereas lexical complexity was construed in terms of lexical density, sophistication, and variation. In general, the results indicated predominantly supportive relationships between fluency and different measures of syntactic complexity but competitive or dual relationships between fluency and lexical complexity. However, the relationships between the selected variables fluctuated over time and often differed in the case of a good, average, and poor language learner.

Keywords: Complex Dynamic Systems Theory, syntactic and lexical complexity, fluency, L2 speech, secondary school

1. Introduction

Complex Dynamic Systems Theory (CDST) constitutes one of alternative approaches to second language acquisition (Atkinson 2011). Following de Bot (2017), the name is used here to refer to both Complexity Theory (CT) (Larsen-Freeman and Cameron 2008) and Dynamic Systems Theory (DST) (Verspoor, de Bot and Lowie 2011) which were developed in different academic centres but share common linguistic and methodological assumptions as well as offer practical instruments to study second language development (SLD). Rooted in a general theory of change, CDST advocates the analysis of changes that take place within the dynamics of this emergent, variable, and self-organising process. The main aim of CDST is to "discover when

and how changes take place in the process of development, how different subsystems develop and interact, and how different learners may have different developmental patterns" (van Dijk, Verspoor and Lowie 2011: 59–60). One of the main assumptions of Complex Dynamic Systems Theory (CDST) is that internally complex language subsystems develop non-linearly while entering different kinds of relationships which may be supportive, competitive, conditional, or dual. What is more, these subsystems remain in the state of competition for learners' limited linguistic and cognitive resources, which leads to trade-offs between them in that progress in the development of one subsystem may cause some regress in the development of the other subsystem. Such trade-offs are likely to characterise the development of complexity, accuracy, and fluency, especially in spontaneous speech in a foreign language. Indeed, many researchers point out that these three aspects of language proficiency are complex and inter-related phenomena which develop in non-linear ways (Michel 2017).

In general, the present case study focused on two constructs, namely intra-individual variability and dynamic relationships. Intra-individual variability, defined as changes in a given language area on repeated measurements within an individual learner (van Geert and van Dijk 2002), was examined in terms of general and specific measures of language development. The general measures referred to complexity, accuracy, and fluency (Rokoszewska 2019a), while the specific measures to syntactic complexity, defined in terms of subordination, coordination, and nominalisation (Rokoszewska 2019b), and lexical complexity, understood in terms of density, sophistication, variation, and frequency (Rokoszewska 2020a). Dynamic relationships were investigated between the general and specific indices (Rokoszewska 2019a, 2019b) as well as between the general measure of accuracy and the specific indices of syntactic and lexical complexity (Rokoszewska 2020b). Thus, the present part of the case study, i.e. the fifth part, will examine the relationships between another general measure, namely fluency and the specific measures of syntactic and lexical complexity.

2. Language fluency and complexity in CDST

Language fluency and complexity form an inherent part of the so-called CALF construct, which in contrast to the so-called CAF triad, refers not only to complexity, accuracy, and fluency, it but differentiates between syntactic and lexical complexity as well. Thus, the acronym CALF stands for syntactic complexity (C), accuracy (A), lexical complexity (L), and fluency (F). Having examined the relationships between these two types of complexity and accuracy, the present part of the case study focuses on such relationships with respect to fluency.

With respect to L2 speech, fluency is defined construed in terms of speed, silence, and repair (Tavakoli and Skehan 2005). Speed or rate is determined by one's access to and control of proceduralized knowledge during language processing. It may be measured in terms of speech rate (SR), i.e. the number of syllables per second, articulation rate (AR), i.e. the number of syllables divided by total speech time excluding corrections, repetitions, false starts, and pauses, and the mean length of run (MLR), i.e. the number of syllables divided by the number of utterances between pauses (Kormos and Denes 2004, Taylor 2018). Silence or breakdown reflects the stages of conceptualisation and planning in language production (Levelt 1989, de Bot 1992). It is usually described by the number, duration, and location of pauses which may

appear in the middle of clauses or at their boundaries. Alternatively, phonation time ratio, i.e. "the percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample", can be computed (Kormos and Denes 2004: 148). Repair reflects pre-articulatory and post-articulatory monitoring and may be measured by the number of false-starts, repetitions, and self-corrections per one hundred words. With respect to L2 writing, fluency may be measured in terms of rate, i.e. the number of words produced per minute calculated on the basis of the final version of the text, or in terms of length, i.e. the number of words per utterance (Michel 2017). It is important to add that the use of keystroke logging software enables the analysis of the writing process in that it is possible to calculate the number of characters written between the pauses or the ratio of characters produced during writing per characters included in the final version of the text (Leijten and van Waes 2013).

Complexity is construed in cognitive terms as "the number of discrete components that a language feature or a language system consists of, and the number of connections between the different components" (Bulte and Housen 2012: 24). It is usually divided into grammatical and lexical complexity. The analysis of grammatical complexity may involve the analysis of syntax, morphology, and phonology with respect to the length of the production unit, e.g. the number of words per clause, sentence or T-unit, the variety of units, e.g. the number of different morphemes, and the interdependence between the units, e.g. coordination vs. subordination (Bulte and Housen 2012; Michel 2017). The analysis of syntactic complexity in a time developmental series should be based on measuring coordination, subordination, and nominalisation since they are good indicators of language complexification at lower, intermediate, and higher levels, respectively (Norris and Ortega 2009). The analysis of lexical complexity or richness may involve the analysis of lexical variation, i.e. the use of different words in a text, lexical density, i.e. the use of lexical items in a text, lexical sophistication, i.e. the use of advanced words, and lexical accuracy, i.e. the types and number of lexical errors (Read 2000).

According to Complex Dynamic Systems Theory (CDST), complexity, accuracy and fluency function as the so-called coupled or connected growers which create supportive, competitive, or conditional relationships in the course of language development (van Dijk *et al.* 2011). Supportive growers develop hand in hand, both either increasing or decreasing. Competitive growers or competitors alternate so that if one variable goes up, the other goes down and vice versa. Conditional growers or precursors develop in such a way that the development of one grower is a pre-condition for another grower to be developed later. Such relationships may be observed in non-linear development of complex language subsystems because these subsystems compete for the learner's limited cognitive and linguistic resources giving rise to trade-offs between them, especially in speech. Such relationships may not be smooth, static, and similar for all learners but fluctuant, dynamic, and different for individual learners (van Dijk *et al.* 2011).

The present case study was divided into several parts. The primary aim of the whole case study was to examine the phenomenon of intra-individual variability in the emergence of general measures of language development, i.e. complexity, accuracy, and fluency, as well as specific measures of syntactic complexity, i.e. subordination, coordination, and nominalisation, and lexical complexity, i.e. lexical density, sophistication, and variation in speaking English as

a foreign language at secondary school in the case of a good, average, and poor language learner. In general, the results indicated that there were no statistically significant differences between the patterns of intra-individual variability in the development of these measures but the relationship between the learners' level of intra-individual variability and the rate of development of these variables was positive (Rokoszewska 2019a, 2019b, 2020a). The secondary aim was to investigate dynamic relationships which take place between the general and specific measures in a time series. So far the relationships between complexity, accuracy, and fluency (Rokoszewska 2019a), including the relationships between specific measures of syntactic complexity (Rokoszewska 2019b), have been analysed. In addition, the influence of specific measures of syntactic and lexical complexity on accuracy has been examined (Rokoszewska 2020b). In general, the results indicated that the development of language subsystems, the trajectories of the learning paths, and the types of relationships between different language variables were non-linear, dynamic, and learner-specific. This analysis, however, needs to be completed by the examination of the relationships between specific measures of syntactic and lexical complexity, on the one hand, and fluency on the other.

3. Research design

The aim of the present part of the case study was to examine the role of lexical and syntactic complexity in the development of fluency in speaking English as a foreign language at secondary school in the case of a good, average, and poor language learner. More precisely, the goal was to identify different types of moving correlations between these variables which might develop as supportive, competitive, pre-conditional, or dual growers in the case of the selected learners. Hence, the following research questions were formulated:

1. What are the results of a good, average, and poor learner in fluency, lexical complexity, and syntactic complexity in the development of L2 English speech at secondary school?
2. What types of relationships are formed between fluency and lexical complexity, i.e. lexical density, sophistication, and variation, in the development of L2 English speech during secondary school in the case of a good, average, and poor language learner?
3. What types of relationships are created between fluency and syntactic complexity, i.e. general syntactic complexity, subordination, coordination, and nominalisation, in the development of L2 English speech at the level of secondary school in the case of a good, average, and poor language learner?

The research method was an exploratory case study which was based on selected data from The Spoken English Developmental Corpus of Polish Learners (SEDCPL). The corpus, which consists of around 2100 recorded interviews, was created on the basis of the study conducted by means of repeated measurements among 106 learners at one of secondary schools in Poland in 2014–2017 (Table 1). Thus, the case study belongs to a long-term quantitative and qualitative research project. It was exploratory in the sense that it investigated selected phenomena within the CDST framework on the basis of speech samples taken from single learners, which will be followed by a statistical study of the whole research sample. In line with the CDST principles, the study provided dense, longitudinal, and individual data (van Dijk *et al.* 2011). It was based

on three mini-corpora which traced the emergence of language in oral production in English as a foreign language in the case of a good, average, and poor language learner throughout secondary school[1]. Each mini-corpus included 21 semi-structured interviews on different topics carried out every month during the whole learning period at secondary school. Thus, the case study was based on the analysis of 63 conversations. The procedure of building the mini-corpora involved interviewing the learners, providing feedback on their speech, preparing verified transcripts of the recorded conversations, and analysing the samples of the learners' speech which were around 200 words long.

Table 1: *The procedure of building The Spoken English Developmental Corpus of Polish Learners (SEDCPL)*

THE PROCEDURE OF BUILDING THE SPOKEN CORPUS OF LEARNER ENGLISH									
DATA	SEMESTER 1					SEMESTER 2			
	Sept	Oct.	Nov.	Dec.	Jan.	Feb.	March	April	May	June
GRADE 1	Org.	Test 1 Fashion	Test 2 Internet	Test 3 Music	Test 4 Education	Winter break	Test 5 Ecology	Test 6 Pets	Test 7 Work	Test 8 Holidays
GRADE 2	Org.	Test 9 Books & films	Test 10 Shopping	Test 11 Friendship	Test 12 Christmas	Winter break	Test 13 Family	Test 14 Health	Test 15 Fame	Test 16 Home & living
GRADE 3	Org.	Test 17 Love	Test 18 TV	Test 19 Crime	Winter break	Test 20 Terrorism	Test 21 Tolerance	End of school-year	Matura exam	-

In the present part of the case study, a number of variables were identified. All variables were operationalised on the basis of the so-called minimal terminal unit (T-unit) defined as an independent clause with all dependent clauses embedded in it (Hunt 1965). This kind of unit, next to AS unit, i.e. Analysis of Speech Unit (Foster, Tonkyn and Wigglesworth 2000), is claimed to be more reliable than a sentence in speech analysis (Larsen-Freeman 2006). However, it was T-unit, as opposed to AS unit, that was selected for the analysis since the whole research project involves not only learner spoken but also written corpus, and T-unit is suitable for the analysis of both speech and writing (Larsen-Freeman 2006). Similarly, language fluency was operationalised in terms of the measure which can be applied to both oral and written production, namely the length-based measure. More precisely, this dependent variable was operationalised as the average number of words per T-unit in a given speech sample (Larsen-Freeman 2006). The scale for this variable was interval. The independent variable referred to syntactic and lexical complexity, the scale being interval. Syntactic complexity was measured in terms of general syntactic complexity, i.e. the number of clauses per T-unit (C/T) (Ellis and Barkhuizen 2006), subordination, i.e. the number of subordinated clauses per T-unit (DC/T) (Lu 2010), coordination, i.e. the number of coordinated phrases per T-unit (CP/T), and nominalisation, i.e. the number of complex nominal phrases per T-unit (CN/T) (Lu 2010). Lexical complexity was measured in terms of lexical density (LD), i.e. the number of lexical words per all words, lexical sophistication (LS), i.e. the number of words beyond the first 2000 words in The British National Corpus (BNC) per all words, and lexical variation (LV), i.e. a complex ratio of types to tokens (CTTR) which takes into account the length of the text (Ellis

---

[1]   At the time of the research project, secondary school in Poland included 3 grades consisting of learners at the age of 16–19. Since the 1st of September 2019 it will include 4 grades consisting of learners at the age 15–18.

and Barkhuizen 2005; Larsen-Freeman 2006). The intervening variable, expressed on the interval scale, was construed as the impact of language complexity on the development of fluency in speaking English as a foreign language at secondary school. The moderator variable, i.e. learners' age, was established on the basis of the nominal scale. The control variables, expressed on the basis of the nominal scale, referred to the same nationality, student's book, number of English classes per week as well as no longer visit in the target-language country.

The data were analysed by such computer programmes as Syntactic Complexity Analyser (Lu 2010) and Lexical Complexity Analyser (Ai and Lu 2010; Lu, 2012). In addition, some CDST procedures (Verspoor, Lowie, van Geert, van Dijk and Schmid 2011) were used to examine the so-called moving correlations which illustrate how the relationship between selected variables developed over a longer period of time. The correlations were calculated on the basis of normalised and detrended data and plotted by means of the so-called moving window of correlations in which each measurement point takes into account the previous measurement point.

The subjects in the present case study were 16-year-old secondary school learners who followed an extended programme with 4–6 English lessons per week. The subjects were selected as representatives of good, average, and poor learners on the basis of the results they obtained for three assignments: the placement test, a written essay, and an oral interview. The good learner (GL) gained 5.5 points, the average learner (AL) – 3.45 points and the poor learner (PL) – 2.17. The learners' family background and learning results are presented in Table 2.

Table 2: *The subjects in the present study*

DATA	GOOD LEARNER			AVERAGE LEARNER			POOR LEARNER		
GENDER	female			male			male		
AGE	16–19 (grades 1–3)								
EXPOSURE TO L2	10 years (grade 1); 4–6 lessons (1–3 grades) – extended English programme								
	no extra classes, no longer stay in an L2 country								
RESIDENCE	city			village			city		
EDUCATION (F/M)[2]	higher / higher			secondary / higher			higher / higher		
EMPLOYMENT (F/M)	white collar worker /			blue collar worker /			white collar worker /		
	white collar worker			white collar worker			white collar worker		
ENGLISH (F/M)[3]	very good / basic			basic / average			very good / basic		
GPA	5.01			4.25			3.54		
GRADES IN ENG.	5.17			3.92			2.67		
FINAL EXAM (%)	Basic	Extended	Oral	Basic	Extended	Oral	Basic	Extended	Oral
	100.0	98.0	100.0	70.0	66.0	77.0	98.0	-	96.0
CLASSIFICATION (pts./ grades)	Test	Speak.	Writ.	Test	Speak.	Writ.	Test	Speak.	Writ.
	6.0	5.0	5.5	3.0	3.75	3.5	1.0	2.0	3.5
	(93pts.)			(61pts.)			(36pts.)		
	Total – 5.5 pts.			Total – 3.42 pts.			Total – 2.17 pts.		

---

[2]   F/M – father/ mother

[3]   The students' opinions about their parents' knowledge of English.

4. Research results

4.1. *Fluency and complexity – general results*

With respect to language fluency (Table 3), it was found out that the good learner, on average, produced 11.30 (SD=2.10), the average learner – 10.10 (SD=1.79), and the poor learner – 9.45 (SD=2.22) words per T-unit in speaking English as a foreign language at the level of secondary school. The comparison of the learners' scores carried out with the use of one-way ANOVA (p=0.05) and Tukey-Kramer Test, i.e. a means differentiation test, showed that only the difference between the good and poor learner was statistically significant.

With respect to lexical complexity (Table 3), it was established that the good learner, on average, produced around 46.0% (SD=0.04) of lexical tokens and 18.0% (SD=0.07) of sophisticated lexical tokens per all tokens in a speech on a given topic, with the score for the varied use of tokens being equal to 4.40 (SD=0.39). The average learner used 48.0% (SD=0.04) of lexical items, 18.0% (SD=0.06) of which were sophisticated, while his score on lexical variation was 4.04. The poor learner obtained 48.0% (SD=0.06) for lexical density, 21.0% (SD=0.07) for lexical sophistication, and 3.91 (SD=0.33) for lexical variation. In general, the differences between the learners' scores (Table 3) were statistically significant only in lexical variation, except the difference between the average and poor learner.

Table 3: *The development of fluency and lexical complexity – average results*

THE DEVELOPMENT OF FLUENCY AND LEXICAL COMPLEXITY – AVERAGE RESULTS											
DATA	FLUENCY			LEX. DENSITY			LEX. SOPHISTICATION			LEX. VARIATION		
	GL	AL	PL	GL	AL	PL	GL	AL	PL	GL	AL	PL
MEAN	11.30	10.10	9.45	0.46	0.48	0.48	0.18	0.18	0.21	4.40	4.04	3.91
SD	2.10	1.79	2.22	0.04	0.04	0.06	0.07	0.06	0.07	0.39	0.47	0.33
MIN	8.42	7.19	6.34	0.12	0.40	0.36	0.03	0.10	0.12	3.76	3.14	3.27
MAX	14.29	13.71	14.20	0.35	0.56	0.60	0.31	0.26	0.35	5.02	5.15	4.41
ANOVA	0.018			0.505			0.505			0.001		
TUKEY-KRAMER TEST[4]	GL=AL GL≠PL AL=PL			-			-			GL≠AL GL≠PL AL=PL		

With respect to syntactic complexity (Table 4), it was found out that the good learner, on average, produced 2.30 (SD=0.92) clauses, 1.00 (SD=0.61) subordinated clauses, 0.40 (SD=0.33) coordinated phrases, and 1.70 (SD=1.00) complex nominal phrases per T-unit in speaking English at secondary school. The average learner used 1.47 (SD=0.21) clauses, 0.47 (SD=0.19) subordinated clauses, 0.24 (SD=0.14) coordinated phrases, and 0.80 (SD=0.20) complex nominals per T-unit. The poor learner built 1.51 (SD=0.35) clauses, 0.50 (SD=0.27) subordinated clauses, 0.32 (SD=0.11) coordinated phrases, and 0.86 (SD=0.39) complex nominal phrases per T-unit in speaking English at secondary school. The results of the statistical analysis conducted by means of one-way ANOVA (p=0.05) indicated that the differences between the three learners were statistically significant in all measures of syntactic complexity, except coordination (Table 4). However, Tukey-Kramer Test revealed that these differences

---

[4] As this test involves the comparison of absolute difference and critical range, detailed numbers are not provided here.

were significant only between the good and average learner as well as between the good and poor learner (Table 4).

Table 4: The development of syntactic complexity – average results

THE DEVELOPMENT OF SYNTACTIC COMPLEXITY – AVERAGE RESULTS											
DATA	GENERAL SYNTACTIC COMPLEXITY			SUBORDINATION			COORDINATION			NOMINALISATION		
	GL	AL	PL	GL	AL	PL	GL	AL	PL	GL	AL	PL
MEAN	2.30	1.47	1.51	1.00	0.47	0.50	0.40	0.24	0.32	1.70	0.80	0.86
SD	0.92	0.21	0.35	0.61	0.19	0.27	0.33	0.14	0.11	1.00	0.29	0.39
MIN	1.15	1.04	0.90	0.21	0.07	0.07	0.15	0.00	0.16	0.25	0.43	0.19
MAX	4.67	1.87	2.13	2.75	0.93	1.07	1.23	0.53	0.53	4.17	1.67	1.65
ANOVA	0.000			0.000			0.163			0.000		
TUKEY-KRAMER TEST	GL≠AL GL≠PL AL=PL			GL≠AL GL≠PL AL=PL			-			GL≠AL GL≠PL AL=PL		

4.2. Moving correlations between fluency and lexical complexity

Analysing the relationships between fluency and lexical complexity in a time series, it was observed that the correlation between fluency and the first lexical measure, i.e. lexical density, was very weak and negative for the good (−.3130) and poor (−.2192) learner but non-existent for the average learner (−.0166) (Table 5). In the case of the good learner, moving correlation indicated that the relationship between the two variables was predominantly negative, especially in the second part of the observation period (Figure 1). In the case of the average learner, the relationship was pre-conditional in that the two variables first competed but then mainly supported each other (Figure 1). In the case of the poor learner, the relationship was dual, with high competition between the variables in the middle of the learning period and very low support at the beginning and end of this period (Figure 1).

Table 5: Correlations and relationships between accuracy and lexical complexity measures

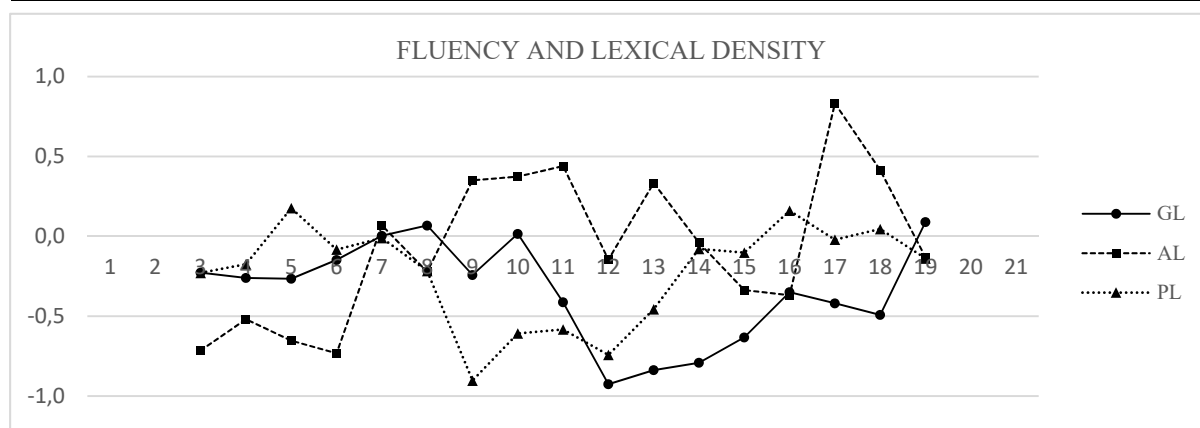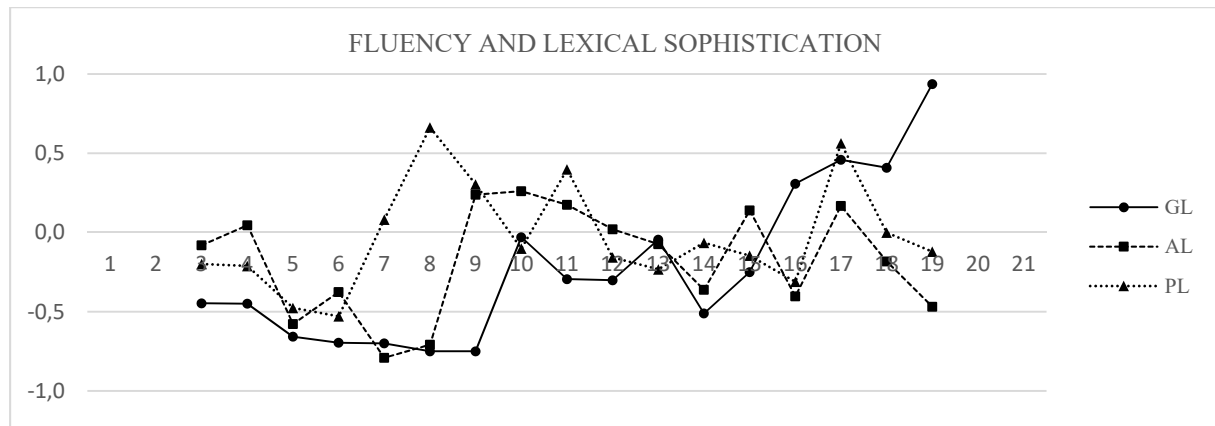| ACCURACY AND LEXICAL COMPLEXITY – CORRELATIONS AND RELATIONSHIPS							
DATA	Ls	LEX. DENSISTY		LEX. SOPHISTICATION		LEX. VARIATION	
		COR.	REL.	COR.	REL.	COR.	REL.
FLUENCY	GL	−0.3130	comp.	−0.2474	pre-cond.	−0.1538	dual
	AL	−0.0166	pre-cond.	−0.0626	dual	0.0717	dual
	PL	−0.2192	dual	−0.0161	dual	0.2306	dual

Figure 1: Moving correlations between fluency and lexical density – all learners

The relationship between fluency and the second measure of lexical complexity, i.e. lexical sophistication, was very weak and negative for the good learner (−.2474) but non-existent for the average (−.0626) and poor learner (−.0161) (Table 5). In the case of the good learner, moving correlation illustrated a pre-conditional relationship (Figure 2). In the case of the average and poor learner, the relationship was dual in that the two variables alternated in moderate support and competition (Figure 2).

Figure 2: *Moving correlations between fluency and lexical sophistication – all learners*

The relationship between fluency and the third measure of lexical complexity, i.e. lexical variation, was weak and positive for the poor learner (.2306) but non-existent for the good (−.1538) and average learner (.0717) (Table 5). In terms of moving correlations, the relationship between the two factors was best described as dual in the case of all three learners (Figure 3) in that the variables functioned as intermittent supporters and competitors.
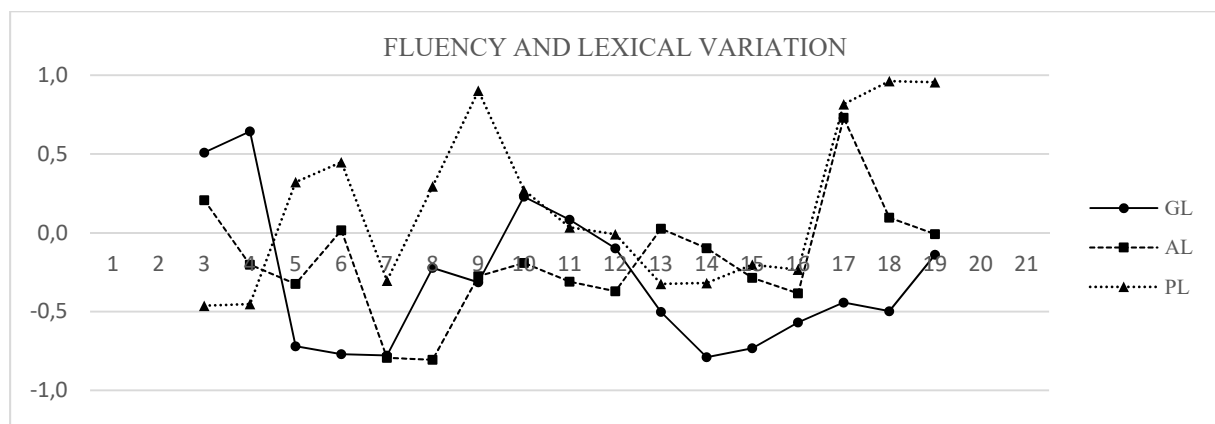
Figure 3: *Moving correlations between fluency and lexical variation – all learners*

4.3. *Moving correlations between fluency and syntactic complexity*

Analysing the relationships between fluency and syntactic complexity in a time series, it was established that the correlation between fluency and general syntactic complexity indicated a weak, positive relationship for the good learner (.3742) and a strong, positive relationship for the average (.8881) and poor (.8238) learner (Table 6). This was confirmed by the so-called moving correlations which illustrated how the relationship between the two variables changed

over the period of three years. In other words, these correlations indicated that fluency and general syntactic complexity developed as the so-called supportive growers in the case of all three learners, though in the case of the good learner some decrease in this support could be observed between data collection points 12–16 (Figure 4).

Table 6: *Correlations and relationships between fluency and syntactic complexity measures*

FLUENCY AND SYNTACTIC COMPLEXITY – CORRELATIONS AND RELATIONSHIPS									
DATA	Ls	GEN. SYNTACTIC COMPLEXITY		SUBORDINATION		COORDINATION		NOMINALISATION	
		COR.	REL.	COR.	REL.	COR.	REL.	COR.	REL.
FLUENCY	GL	0.3742	supportive	0.5999	supportive	0.4607	supportive	0.4610	supportive
	AL	0.8881	supportive	0.9168	supportive	0.5497	supportive	0.6935	supportive
	PL	0.8238	supportive	0.8420	supportive	0.2551	supportive	0.4653	supportive
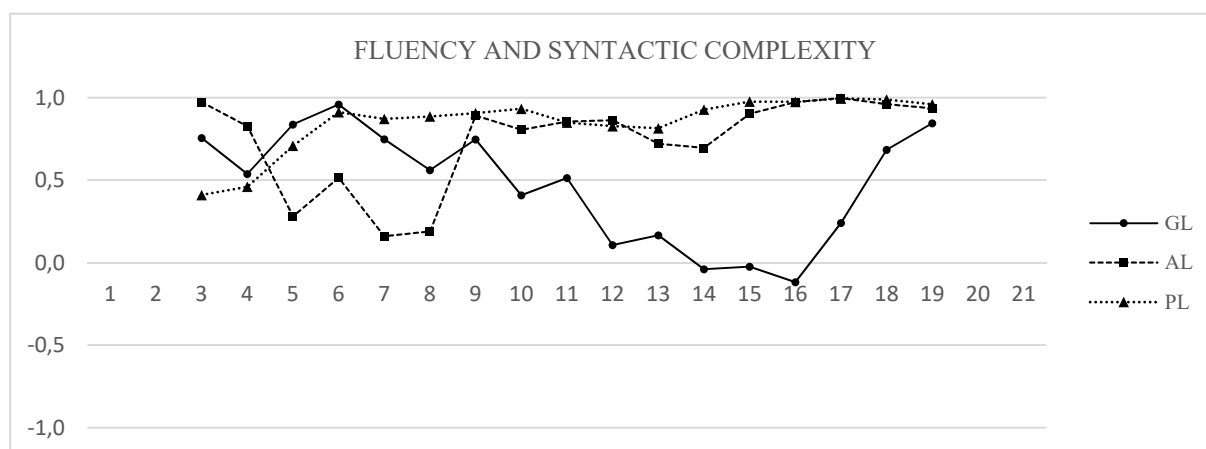
Figure 4: *Moving correlations between fluency and general syntactic complexity – all learners*

With respect to more specific measures of syntactic complexity (Table 6), it was observed that there was a positive relationship between fluency and subordination in the case of all three learners. This relationship was very strong for the average (.9168) and poor (.8420) learner but moderate for the good learner (.5999). In congruence with these findings, moving correlations revealed consistent patterns of mutual support between the two variables in a time series in the case of the average and poor learner. In the case of the good learner, the pattern was less stable as it involved some decrease in the supportive relationship between data points 12–16 (Figure 5).
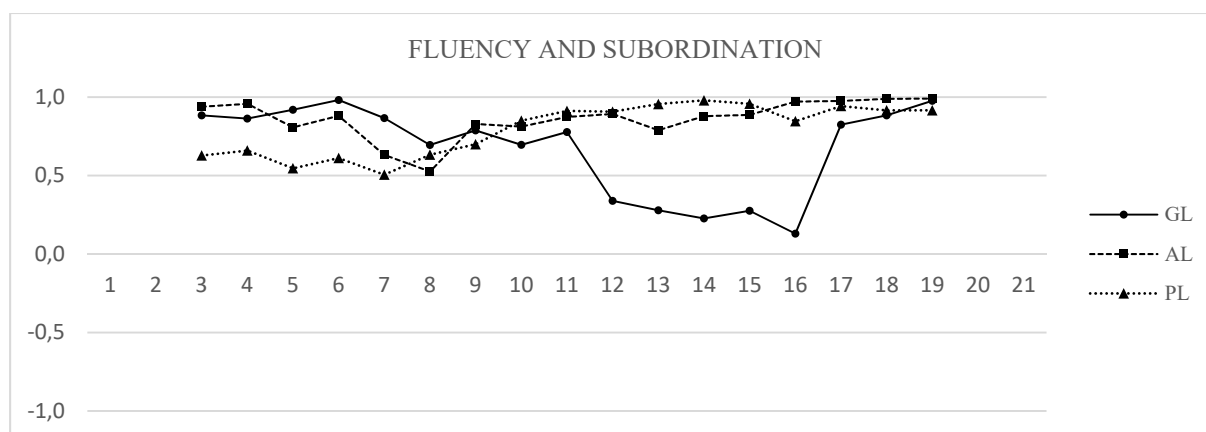

Figure 5: *Moving correlations between fluency and subordination – all learners*

With respect to fluency and phrasal coordination, standard correlations indicated a positive relationship between the two variables, which was moderate for the good (.4607) and average (.5497) learner but weak for the poor learner (.2551) (Table 6). The visual analysis of moving correlations confirmed that the relationship in question was supportive for the good and average learner (Figure 6). However, in the case of the former, some pre-conditioning could be observed at the beginning of the observation period. In the case of the latter, support dropped twice during the observation period, namely between data points 5–7 and 16–17, which indicated some duality in the course of development. Furthermore, in the case of the poor learner, the relationship was clearly pre-conditional because first it was predominantly competitive, except two outliers at the beginning, and then it became more supportive at the end (Figure 6).
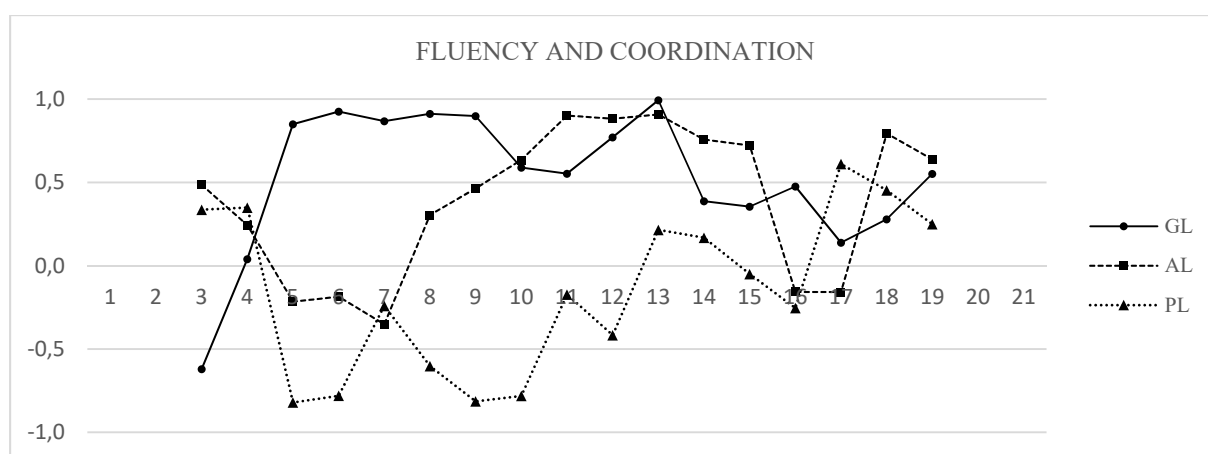

Figure 6: *Moving correlations between fluency and coordination – all learners*

With respect to fluency and nominalisation, it was found out that the relationship between the two factors was moderate and positive for all three learners (GL – .4610; AL – .6935; PL – 4653) (Table 6). However, moving correlations indicated that this support was not fixed and stable. In the case of the good learner, such support, high at the beginning of the observation period, decreased in the second part of this period but became stronger at the end (Figure 7). In the case of the average learner, such support dropped twice (data points 5–7 and 14–15), whereas, in the case of the poor learner, it went down once (data points 11–12) (Figure 7).
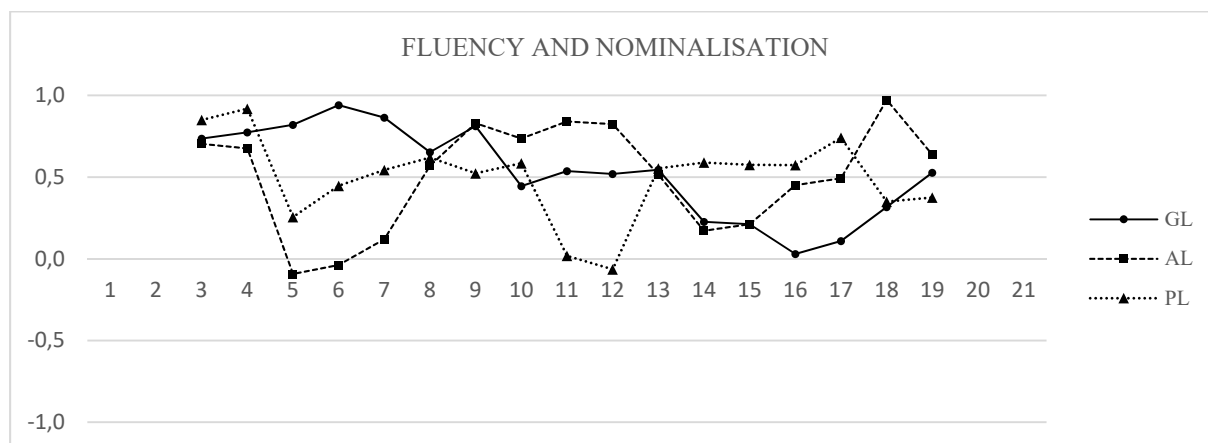

Figure 7: *Moving correlations between fluency and nominalisation – all learners*

5. Discussion

The aim of the present paper, which presents the fifth part of the case study, was to explore the influence of lexical and syntactic complexity on the development of fluency in speaking English as a foreign language at secondary school on the example of a good, average, and poor language learner. With respect to the first research question, which referred to the learners' results on fluency and complexity in L2 English speech, it was found out that the good learner's speech was more fluent than the speech of the poor but not average learner (Rokoszewska 2019a). The good learner's speech was also more syntactically complex than the speech of the average and poor learner in terms of all syntactic measures, such as general sentence complexity, subordination, and nominalisation, but not phrasal coordination (Rokoszewska 2019b). What is more, this learner's speech was more lexically complex than the speech of the other two learners only in terms of lexical variation as opposed to density and sophistication (Rokoszewska 2020a).

With respect to the second research question, which focused on the relationships between fluency and lexical complexity in L2 English speech, it was observed that these relationships were characterised with greater competition. Most of the relationships were dual in that fluency, on the one hand, and lexical density, sophistication, and variation on the other hand, developed as intermittent supporters and competitors. Fluctuations in the trajectories of moving correlations between fluency and different measures of lexical complexity might have been related to the learners' ability to use lexis connected with a given topic. It may be assumed that if learners managed to access denser, more sophisticated, and more varied lexis easily in written production, then the length of the production was supported. Conversely, if they struggled to recall such lexis, the length of the production unit was compromised.

With respect to the third research question, which concerned the relationships between fluency and syntactic complexity in L2 English speech, it was established that the relationships between syntactic complexity and fluency were generally positive in the case of all three learners. It may be concluded that fluency and different measures of syntactic complexity developed as the so-called connected supportive growers, which means that the use of complex sentences, subordinated clauses, coordinated phrases, and complex nominal phrases contributed to the length of the production unit, i.e. T-unit, in speech. However, the support between fluency and different measures of syntactic complexity was not constant but changeable over the whole learning period. Lower support in the trajectories of moving relationships between fluency and different measures of syntactic complexity, in particular coordination and nominalisation, may indicate that the learners used coordinated and nominal phrases in simple and coordinated sentences, which affected the length of the T-unit, as producing such phrases in complex subordinated sentences was linguistically and cognitively more demanding.

In addition, it may be also observed that the relationship between fluency and lexical complexity depends to some extent on the relationship between fluency and syntactic complexity. The example of the good learner indicated that the relationship between fluency and different syntactic measures, such as subordination, coordination, and nominalisation, was high in the first part of the observation period, but it became weaker when the pre-conditional relationship between fluency and sophisticated words became positive in the second part of this

period. In other words, the use of sophisticated words led to some trade-off with syntactic complexity. Generally, in the case of the good learner, it may be said that fluency was first supported by syntactic and later by lexical complexity. In the case of the average and poor learner, the patterns were more chaotic.

The present case study provided some insight into language development in oral production in the case of individual learners who differed in terms of success at learning a foreign language at secondary school. Since the study was exploratory in character, it focused only on single representatives of good, average, and poor learners. Hence, it would be useful to examine groups of learners who achieve different success at learning a foreign language in the formal context. Another limitation is that fluency was evaluated by means of the length-based measure which could be used for both speech and writing. Still, however, it is possible to recommend that teachers should focus not only on fluency but also syntactic and lexical complexity of learners' speech in terms of instruction and evaluation, recognizing that the development of language complexity, next to accuracy and fluency, is a complex process and a challenging task.

6. Conclusions

Summing up, it is crucial to highlight the fact that Complex Dynamic Systems Theory is one of alternative approaches to applied linguistics which provides new theoretical principles and methodological procedures to study second or foreign language development. On the basis of the present part of the case study, conducted within the CDST framework, a few conclusions were drawn. First of all, it was observed that fluency, syntactic complexity, and lexical complexity emerged as the so-called coupled or connected growers in various types of relationships in a time developmental series. Second of all, the examined relationships were dynamic and fluctuant, which reflected the complexity, non-linearity, and variability of language development. Third of all, the trajectories of moving correlations between fluency and complexity illustrated substantial support in terms of syntax but equally substantial duality of intermittent support and competition in terms of lexis. Such duality reflected trade-offs between these subsystems which were due to learners' constrained language processing, especially in L2 speech. Last of all, such relationships were characteristic for individual learners as the trajectories of these relationships did not overlap. Nevertheless, it should be pointed out that the above conclusions refer to individual learners selected for the purpose of the present case study. In spite of the fact that CDST researchers advocate they study of individual language development, it is necessary to verify these findings with respect to the whole group of learners and/or the groups representing a given type of learners.

References

Ai, H., and X. Lu. 2010, A web-based system for automatic measurement of lexical complexity. Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10). Amherst, MA. June 8–12.

Atkinson D. (ed.), 2011. *Alternative approaches to second language acquisition*. New York, NY: Routledge.

Bulte, B., and A. Housen. 2012. Defining and operationalizing L2 complexity. In A. Hounsen, F. Kuiken, I. Vedder (eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 23–46. Amsterdam/Philadelphia: John Benjamins. Retrieved from https://www.academia.edu/27458664/.

de Bot, K. 1992. A bilingual production model: Level's "Speaking" model adapted. *Applied Linguistics* 13: 1–24. Retrieved from https://www.researchgate.net/profile/Kees_De_Bot2/publication/249237191.

de Bot, K. 2017. Complexity theory and dynamic systems theory: Same or different? In L. Ortega, and Z. Han (eds.), *Complexity theory and language development. In celebration of Diane Larsen-Freeman*, 51–58. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/lllt.48.03deb.

Ellis, R., and G. Barkhuizen, 2005. *Analysing learner language*. Oxford: Oxford University Press.

Foster, P., A. Tonkyn, and G. Wigglesworth. 2000. Measuring a spoken language: A unit for all reasons. *Applied Linguistics* 21(3): 354–375. https://doi.org/10.1093/applin/21.3.354.

Hunt, K. 1965. *Grammatical structures written at three grade levels. NCTE Research report*, *No. 3.* Champaign, IL, USA: NCTE.

Kormos, J., and M. Denes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32: 145–164. Retrieved from https://www.researchgate.net/publication/222670790

Larsen-Freeman, D. 2006. The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* 27: 590–616. Retrieved from https://www.academia.edu/38727423/.

Larsen-Freeman, D., and L. Cameron. 2008. *Complex systems and applied linguistics.* Oxford: Oxford University Press.

Leijten, M., and L. van Waes. 2013. Keystroke logging in writing research: Using InputLog to analyse and visualize writing processes. *Written Communication* 30(3): 358–392. https://doi.org/10.1177/0741088313491692.

Levelt, W. 1989. *Speaking: From intention to articulation*. Cambridge: Cambridge University Press.

Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4): 474–496. Retrieved from https://www.researchgate.net/publication/233686598.

Lu, X. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96(2): 190–208. Retrieved from https://www.researchgate.net/publication/262084862

Michel, M. 2017. Complexity, accuracy and fluency in L2 production. In S. Loewen, and M. Sato (eds.), *The Routledge handbook of instructed second language acquisition*, 50–68. London/New York: Routledge.

Norris, J. M., and L. Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30(4): 555–578. Retrieved from https://www.researchgate.net/publication/249237888.

Read, J. 2000. *Assessing vocabulary*. Oxford: Oxford University Press.

Rokoszewska, K. 2019a. Intra-individual variability in the emergence of complexity, accuracy and fluency in speaking English at secondary school – A case study of a good, average and poor language learner. *Anglica Wratislaviensia* LVII: 181–204.

Rokoszewska, K. 2019b. Intra-individual variability in the emergence of syntactic complexity in English L2 speech at secondary school – A case study of a good, average and poor language learner. *Konin Language Studies* 7(4): 445–473.

Rokoszewska, K. 2020a. Intra-individual variability in the emergence of lexical complexity in speaking English at secondary school – A case study of a good, average and poor language learner. *Theory and Practice in Second Language Acquisition* 6(1): 107–142.

Rokoszewska, K. 2020b. Accuracy and complexity as connected growers in L2 English speech at secondary school – A case study of a good, average and poor language learner. *Neofilolog*: 54(4).

Tavakoli, P., and P. Skehan. 2005. Strategic planning, task structure and performance testing. In R. Ellis (ed.), *Planning and task performance in a second language*, 239–277. Amsterdam/Philadelphia: John Benjamins.

Taylor, S. 2018. Correlating speech fluency measures with speech fluency perceptions. *Language Education and Research Center Journal* 13, 5–21. Etrieved from https://www.researchgate.net/publication/326058705

van Dijk, M., M. Verspoor, and W. Lowie. 2011. Variability and DST. In M. Verspoor, K. de Bot, and W. Lowie (eds.), *A dynamic approach to second language development*, 55–84. Amsterdam/Philadelphia: John Benjamins.

van Geert, P., and M. van Dijk. 2002. Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behaviour and Development*: 25, 340–375. Retrieved from https://www.paulvangeert.nl/publications_files/vangeert_variability.pdf.

Verspoor, M., de Bot, K., and W. Lowie. 2011. *A dynamic approach to second language development.* Amsterdam: John Benjamins Publishing Company.

Verspoor, M., W. Lowie, P. van Geert, M. van Dijk, and M. S. Schmid. 2011, *How to Sections.* In M. Verspoor, K. de Bot, and W. Lowie (eds.), *A dynamic approach to second language development*, 129–199. Amsterdam/Philadelphia: John Benjamins.

Structure constraints in Polish and English adjectival synthetic compounds[*]

Sebastian Wasak

John Paul II Catholic University of Lublin, Poland

Abstract

The subject matter of this paper is the external syntax of adjectival synthetic compounds in Polish (e.g. *czasochłonny*, *ciepłolubny*, *opiniotwórczy*, etc.) and English (*life-giving*, *sleep-inducing*, *far-reaching*, etc.). The primary objective of the study is to determine whether *-ny/-czy/-ły* compounds in Polish and adjectival *-ing* compounds in English, whose heads appear to be derived from verbs, are deverbal in the sense of Distributed Morphology; that is, whether their external syntax points to the presence of complex verbal structure in their syntactic representation. It is shown that adjectival synthetic compounds in Polish and English behave in a way typical of underived adjectives, being unrestricted in the predicative position and allowing degree modification with *very*; as such they are not deverbal in the morphosyntactic sense with their syntactic representation lacking the functional heads vP and VoiceP found in deverbal structures. The limited productivity of adjectival synthetic compounds further contributes to their non-eventive status.

Keywords: adjective, synthetic compound, verbal structure, Distributed Morphology

1. Introduction

The licensing of argument structure by deverbal adjectives has been a widely studies topic by language researchers. In particular, the implicit presence of the external argument in adjectival participles (Kratzer 2000; Embick 2003, 2004; Gehrke 2011; McIntyre 2013; Alexiadou *et al.* 2014, Bruening 2014) has generated a great deal of discussion. This paper will be centred on the question of argument structure in adjectival synthetic in Polish (e.g. *czasochłonny* 'time-consuming', *życiodajny* 'life-giving', *opiniotwórczy* lit. 'opinion-forming, influential', etc.) and their English counterparts (e.g. *time-consuming*, *life-giving*, *sleep-inducing*, etc.). Specifically, the objective of the study is to determine whether the morphosyntactic properties of Polish and English adjectival synthetic compounds point to the presence of argument structure, specifically the external argument, and how their morphosyntactic characteristics are manifested in their internal structure.

The present study is couched within the Distributed Morphology approach, which posits that syntax is a single generative component responsible for the formation of both sentences and words. Furthermore, according to DM, all words are decomposed into roots.

The paper is structured as follows: Section 2 will be devoted to the discussion of what is required of a construction to label it deverbal. In Section 3 we will investigate the external syntax of adjectival synthetic compounds in Polish and English paying special attention to whether their syntactic behaviour points to their adjectival or verbal nature. In this section we will also look into the question of whether adjectival synthetic compounds have the capacity to license the external argument. The syntactic analysis of adjectival synthetic compounds will allow us to propose their internal structure. Our findings will be summarised in Section 4.

2. Deverbal vs. non-deverbal constructions in Distributed Morphology

In the lexicalist tradition, the labels 'deverbal adjective' and 'deverbal noun' have been applied to every adjective and noun derived from a lexical verb by an affix. Consider *scratcher* understood as a lottery ticket which is scratched. At a first glance, *scratcher* appears to be derived from the verb *scratch* by attaching the suffix *-er* to the verbal base.

The DM approach looks at the issue of what is and what is not deverbal differently. For a distributed morphologist, For a distributed morphologist, it is not sufficient to claim that an adjective (or a noun) being based on what appears to be a verbal root is deverbal as DM takes also into account the internal semantics of words. What this means in practice is that deverbal formations are those whose meanings are fully transparent: this is not the case with, for example, *scratcher* where the suffix *-er* is not employed to form an agentive nominal, which is the most common function of the suffix. Thus, the DM approach does not consider nominals of the *scratcher* type[1] to be deverbal – they are not productive and, consequently, convey idiosyncratic meaning (Alexiadou and Schäfer 2010).

DM also attaches crucial importance to the external syntax of complex words. One of the features of productive, semantically predictable *-er* nominals is that they can be preceded by adjectives modifying the events denoted by verbs on which they are based. Such modification is hardly possible with unproductive, idiosyncratic *-er* nominals.[2]

(1) a. *a skilful worker* (= someone who works skilfully)
  b. *an ardent admirer of Picasso* (= someone who admires Picasso ardently)
  c. ??*an easy scratcher* (= a lottery ticket that scratches easily)

Consider also, for example, participles in Greek which come in two variants: one based on the suffix *-menos* and the other incorporating the suffix *-tos*. One of the main differences between the two is that negating prefixes can attach only *-tos* participles (Anagnostopoulou 2003: 10):

---

[1] Other non-deverbal *-er* nominals include *baker*, *broiler* or *diner* (Alexiadou and Schäfer 2010).
[2] As noted by Alexiadou and Schäfer (2010), event adjectives can also be licensed by non-deverbal nominals which are associated with certain actions, as in *a fast horse*.

(2)   a.   *gra-menos*   'written'
        b.   *a-graf-tos*   'unwritten'

(3)   a.   *pli-menos*   'washed'
        b.   *a-pli-tos*   'unwashed'

(4)   a.   *diavas-menos*   'read'
        b.   *a-diavas-tos*   'unread'

(5)   a.   *fago-menos*   'eaten'
        b.   *a-fago-tos*   'uneaten'

(6)   Alexiadou and Anagnostopoulou (2008: 33)
        a.   **adiavasmenos*   'unread'
        b.   **aplimenos*   'unwashed'

The *-menos* and *-tos* participles also differ in that the former, unlike the latter, has the ability to license *by*-phrases (Anagnostopoulou 2003: 11):

(7)   a.   *To   vivlio ine gra-meno apo tin Maria.*
           The   book is   written   by   the Mary
           'The book is written by Mary.'
        b.   **To vivlio ine a-graf-to apo tin Maria.*
           The   book is   unwritten by   the Mary
           'The book is unwritten by Mary.'

The implication of the data above is that only the *-menos* participle has the structure in which the verbalizing head vP and VoiceP (the head introducing the external argument) are projected: it is thus deverbal. On the other hand, the *-tos* participle is derived directly from the root: its structure does not contain little v or VoiceP (Anagnostopoulou and Samioti 2008: 105):

(8)   a.   *-tos*   b.   *-menos*
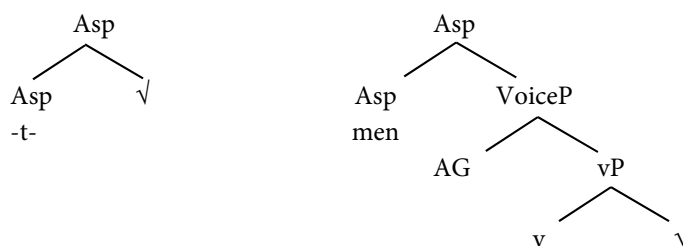
The structure of *-tos* participles is then identical to the structure of simple, underived adjectives, such as *empty* or *dry* (see Embick 2003: 152), the only difference being the presence of the adjectivizing suffix in the Greek participles. The non-eventive external syntax of *-tos* participles means that, despite being morphologically complex, they are simple structurally.

3. Adjectival synthetic compounds

3.1. *Polish* -ny/-czy/-ły *compounds*

In Polish, verbs can be combined with nouns and adverbs to give rise to compound adjectives (Szymanek 2010). Some examples are listed below:

(9) | Noun/Adverb | Verb | Compound |
|---|---|---|
| *czas* 'time' | *chłonąć* 'absorb' | *czasochłonny* 'time-consuming' |
| *mięso* 'meat' | *żreć* 'eat' (obsolete) | *mięsożerny* 'carnivorous' |
| *ropa* 'oil' | *nosić* 'bear' | *roponośny* 'oil-bearing' |
| *kwas* 'acid' | *odpierać* 'resist' | *kwasoodporny* 'acid-resistant' |
| *lekko* 'lightly' | *strawić* 'digest' | *lekkostrawny* 'light, easily digestible' |
| *ciepło* 'warmth' | *lubić* 'like' | *ciepłolubny* 'stenothermic' |
| *dźwięk* 'sound' | *naśladować* 'imitate' | *dźwiękonaśladowczy* 'onomatopoeic' |
| *rak* 'cancer' | *tworzyć* 'form' | *rakotwórczy* 'carcinogenic' |
| *opinia* 'opinion' | *tworzyć* 'form' | *opiniotwórczy* 'lit.opinion-forming, influential' |
| *długo* 'long' | *trwać* 'last' | *długotrwały* 'long, long-lasting' |

The external syntax of *-ny/-czy/ły* compounds in Polish is stative as they exhibit a range of features typical of adjectives. For example, they can appear in the predicative position after a copular verb and be coordinated with other adjectives and adjectival compounds[3] (all examples have been extracted from the National Corpus of Polish, except where noted otherwise):

(10)   a.   Zabezpieczenie hipoteczne wymaga bowiem założenia dla nieruchomości księgi wieczystej, co jest **czasochłonne i kosztowne**.
'Securing mortgage requires the establishment of a land and mortgage register, which is **time-consuming and expensive**.'

b.   Jako substancja chemiczna mazut wykazuje właściwości **rakotwórcze i toksyczne**.
'As a chemical substance, mazout has **carcinogenic and toxic** properties.'

c.   Największe zagrożenie dla upraw roślinnych w okresie zimowym mogą stanowić: silne mrozy przy braku lub niedostatecznej grubości pokrywy śnieżniej, **długotrwała i zbyt gruba** pokrywa śnieżna (…).
'The greatest threat to vegetable crops in winter can be: strong frosts in the absence of or insufficient thickness of the snow cover, **long lasting and too heavy** snow cover (…).'

d.   Szparagi są **lekkostrawne i niskokaloryczne**.
'Asparagus is easily digestible and low in calories.'

e.   Jeżeli ma być to organ **opiniotwórczy i doradczy**, to – w moim przekonaniu – wystarczyłoby, aby powoływał go minister do spraw gospodarki.
'If it is to be an **opinion-forming and advisory** body, then – in my opinion – it would suffice for it to be appointed by the minister for economy.'

f.   Gdyby istotnie przodek człowieka był zwierzęciem **mięsożernym i drapieżnym**, to trzeba przyznać, że byłby to absolutny ewenement "dentystyczny".
'If the human ancestor were a **predatory and carnivorous** animal, it must be admitted that it would be an absolute "dental" revelation.'

---

3   The ability to be coordinated with other adjectives and to appear postnominally has been claimed to be indicative of adjectival nature of participles (Levin and Rappaport 1986; McIntyre 2013).

Polish *-ny/-czy/ły* compounds also accept the modification with adjectival degree modifiers such as *very* (a property often considered to be indicative of the non-verbal nature of a participle, e.g. Kennedy and McNally 1999):

(11)   a.   Praca przy stworzeniu pełnej dokumentacji z przykładami jest **bardzo czasochłonna**.
            'Working on creating full documentation with examples is **very time-consuming**.'

       b.   Są to osiedla wybudowane w pewnym okresie czasu, kiedy do budowy używanolepików, w składzie których były **bardzo rakotwórcze** węglowodory aromatyczne.
            'These are settlements built during a certain period of time, when glues containing very carcinogenic aromatic hydrocarbons were were used for construction.'

       c.   Procedura udowodnienia nieprawidłowości jest **bardzo długotrwała**.
            'The procedure of proving irregularities is very long'.

       d.   (…) tego właśnie dnia łódzką operę odwiedzi liczna i **bardzo opiniotwórcza** grupa krytyków muzycznych (…).
            '(…) on that day, the Łódz Opera House will be visited by a large and very influential group of music critics (…).'

       e.   Ryby są **bardziej lekkostrawne** niż mięso.
            'Fish are more easily digested than meat.'

       f.   Nie ma co liczyć na to, że wraz ze wzrostem zamożności staniemy się **bardziej mięsożerni**.
            'There is no reason to hope that as we become richer we will become more carnivorous.

Finally, Polish *-ny/-czy/ły* compounds accept the prefixation with the negating affix,[4] which is another feature claimed to be very adjectival (Wasow 1977, Levin and Rappaport 1986, McIntyre 2013):[5]

(12)   a.   A przecież dania te ze wszech miar zasługują na uwagę. Sporządza się je głównie z naturalnych, łatwo dostępnych składników, a ich wykonanie jest zwykle proste i **nieczasochłonne**.
            'And yet these dishes by all means deserve attention. They are prepared mainly from natural, easily available ingredients, and preparing them is usually simple and **non-time-consuming.**'

       b.   Jakiś procent populacji jest **niemięsożerny**, a inny wprost przeciwnie.
            Some percentage of the population is **not carnivorous**, while the other is just the opposite.'

       c.   Materiały te są poza tym niedrażniące i **nierakotwórcze**.
            'These materials are also non-irritating and **non-carcinogenic**.'

       d.   Kiedyś liczył się głos Pauline Kael z „New Yorkera", ale zrezygnowała ona ze stałej współpracy. Reszta to płotki **nieopiniotwórcze**.
            'Once the voice of Pauline Kael from New Yorker mattered, but she gave up permanent cooperation. The rest are **non-influential** minnows.

       e.   Łączenie ze sobą w ciągu kilku godzin śledzi, mięs, ciast i zdecydowanie **nielekkostrawnych** sałatek daje naszemu układowi pokarmowemu niezłe wyzwanie.[6]
            'Eating herring, meat, cake and **not easily digestible salads** within a few hours is quite a challenge for the digestive system.'

---

[4] The negated variants of Polish *-ny/-czy-ły* compounds are rarely employed due to the existence of viable antonyms. For example, *niedługotrwały* 'non-long-lasting' is a possible word in Polish but virtually unattested since the majority of speakers show clear preference for *krótkotrwały* 'short-lasting'.

[5] As recently argued in the literature (e.g. Bruening 2014), deverbal adjectives equipped with a negating prefix may host complex structure.

[6] The example (12f) from www.hellozdrowie.pl/blog-swieteczne-zasady-na-mniejsze-zlo/.

In theory, *-ny*/*-czy*/*-ły* compounds in Polish are virtually unrestricted in being able to occur in adjectival context, although some *nie-* forms (e.g. *niekwasooodorny* 'non-acid-resistant', *nieciepłolubny* 'non-stenothermic') are rarely attested due to pragmatic reasons. In addition, the relational *dźwiękonaśladowczy* 'onomatopoeic' cannot be modified by *very*, prefixed with *un-* or attested attributively.

3.2. Adjectival -ing *compounds in English*

Adjectival synthetic compounds in English are formed by merging a modifier with the active participle of a verb.[7] As with Polish, the left-hand side element can be a noun or an adverb:

(13)   a.   *time-consuming*
       b.   *life-giving*
       c.   *sleep-inducing*
       d.   *thought-provoking*
       e.   *law-abiding*
       f.   *life-threatening*
       g.   *fast-acting*
       h.   *slow-melting*
       i.   *far-reaching*
       j.   *forward-looking*

In terms of the external syntax, the behaviour of adjectival *-ing* compounds in English is very similar to the behaviour of *-ny*/*-czy*/*-ły* compounds in Polish. They are found predicatively and in coordination with 'pure' adjectives:

(14)   a.   The victim's wound was not believed to **be life-threatening**.
       b.   Examining a department or agency, its personnel, and its implementation policies **is time-consuming**.
       c.   Reading, reflecting on, and debating the writings of philosophers and educators in the field, especially within the context of a well-taught philosophy of music education course, can **be life-changing**.
       d.   Even tree-huggers like to haul boats and trailers from time to time. But the gas required can **be guilt-inducing**.
       e.   (…) the consequences of teacher stress are **far-reaching**.
       f.   Because the geographic origins of the activities are diverse, the topics covered **are wide-ranging**.

(15)   a.   Childhood cancer is a **life-threatening and traumatic** event that affects the patient as well as the entire family.
       b.   The researchers streamlined the pre-test process so that traditional genetic counselling, which can be **time-consuming and difficult**, was excluded.
       c.   The personal challenges faced by the young black journalist are **thought-provoking and compelling**.
       d.   The Irish were not only more **law-abiding and sober** than their English counterparts but also apparently possessed a superior morality as well (…).

---

[7]   Adjectival compounds in English may be also based on simple adjectives (e.g. *oil-rich*, *lead-free*) and passive participles (e.g. *home-made*, *pencil-drawn)*. Compounds of this type lie beyond the scope of this study.

e. The Senate is poised to begin the most **wide-ranging and ambitious** battle over gun control on Capitol Hill in 20 years (...).

f. Contrary to some reports, there have not been **far-reaching and significant** changes to voter ID laws in the battleground states since 2008.

Adjectival *-ing* compounds in English can be modified by *very* and prefixed with a negating affix; both of these properties are also exhibited by Polish *-ny/-czy/-ły* compounds. By way of illustration, consider examples in (16) and (17).

(16)   **very** time-consuming/thought-provoking/life-threatening/far-reaching

(17)   a. In stressful situations eyewitnesses are likely to become more focused on the stimuli that are stress inducing than on **non-stress-inducing** stimuli (...).
    b. Both people in the car suffered **non-life-threatening** injuries.
    c. It would be quite remarkable to hold that speech by a law-abiding possessor of information can be suppressed in order to deter conduct by a **non-law-abiding** third party.
    d. A powerful body of laboratory-based research has established the effects of fear-arousing appeals versus **non-fear-arousing** appeals.
    e. In addition to being edifying, encouraging, and empowering, our multiple book series are also meant to be **non-time consuming.**
    f. Younger Indian children were equally likely to choose the hard-working and the **non-hard-working** child in response to the question about who would do better with effort and which of the two was smarter.

Adjectival compounds of the *time-consuming* type are contrasted with deverbal compounds which involve event implications, e.g. *decision-making*, *gift-giving* or *award-winning*, which can only function as prenominal modifiers:[8]

(18)   a. a decision-making unit / *This unit is decision-making.
    b. a book-reading family / *This family is book-reading.
    c. a gift-giving person / *This person is gift-giving.
    d. a note-taking student / *This student is note-taking.
    e. degree-seeking people / *These people are not degree-seeking.
    f. an award-winning artist / ?This artist is award-winning.[9]

Another important difference between the compounds in (17) and stative *-ing* compounds is that the latter commonly function as nominals (with the exception of compounds such as *award-winning* which express results of prior events). Conversely, the *time-consuming*-type compounds are impossible as nominals:

(19)   a. rational decision making
    b. recreational book reading

---

8 Eventive *-ing* compounds do exhibit certain adjectival features; for example, they can universally be prefixed with *non-* (*non-tax-paying businesses*, *non-decision-amking units*). They are, however, highly restricted with *very* (**very tax-paying/decision-making/note-taking/degree-seeking*).

9 While *award-winning* is possible predicatively, it appears to be highly degraded in that position. The COCA corpus returns 8 occurrences of the predicative uses of *award-winning* versus over 4000 prenominal occurrences.

> c. ceremonial gift giving
> d. detailed note-taking
> e. *the life giving (by the river)
> f. *the time-consuming (by the procedure)

Polish does not have nominal-adjectival deverbal compounds. While nominal compounds denoting activities do exist in Polish (e.g. *grzybobranie* 'mushroom picking'), they do not give rise to adjectives (*grzybobiorczy*).[10]

3.3. *Adjectival synthetic compounds and the external argument.*

The presence of the external argument in deverbal constructionsis most commonly diagnosed by testing their capacity to license *by*-phrases. This method, however, is hardly applicable to -*ny*/-*czy*/-*ły* and -*ing* compounds as they are adjectives based on active participles. Interestingly, English and Polish adjectival synthetic compounds may be preceded with *deliberately*/*celowo*,[11] as shown in the examples below:

(20) a. The options Niño offers to us are **deliberately thought-provoking** both in the individual content and in their scope as a whole.
     b. The evidence of the New Testament and of early Christian literature shows consistently that the Christians were **deliberately law-abiding** (…).
     c. Without exception, everyone at Saydnaya is subjected to a sustained and **deliberately life-threatening** programme of relentless torture.
     d. This is the first of several articles on the subject to run over the next few months. It is **wide-ranging** in content and emotional reach.
     e. This coalition includes government workers (...) and both wings of the "dependency lobby" (those locked into welfare dependency and those who earn a handsome living managing that **deliberately never-ending** dependency).
     f. On the other hand, Monk's **deliberately awkward-sounding** pianism has never been as popular as his compositions.

(21) Jak wiemy, osadzenie prawne PKW w polskiej strukturze władzy jest **celowo wieloznaczne** i nie jest oczywiste to, kto sprawuje tam władzę (...).
'As we know, the legal establishment of PKW in the Polish power structure is **deliberately ambiguous** and it is not obvious who exercises the power there (...).'

---

[10] Polish has a small inventory of compounds ending in -*czy* which have nominal counterparts (e.g. *krwiodawczy* and *krwiodastwo*). Such compounds are of highly relational nature – they appear only as postnominal modifiers and are not possible predicatively:

   (i)   a. *akcja krwiodawcza*
        'a blood donation event'
      b. ** Ta akcja jest krwiodawcza*
        'Ths event is a blood donation'

[11] The licensing of adverbials of the *deliberately* type has been closely linked with the agentive nature of deverbal structures (Embick 2004; Meltzer-Asscher 2011). They are attested, for example, in deverbal nominals in Polish (Cetnarowska 2017: 150):

   (i)   *umyślne prezydenckie ułaskawienie gangsterów*
      '(the) president's pardoning the gangsters intentionally'

However, the above data hardly point to the agentive nature of *-ing* and *-ny/-czy/-ły* compounds because *deliberately* frequently modifies simple adjectives, as in the COCA examples below:

(22) a.   Some members fo the media were engaged in deliberately false reporting.
     b.   This deceptively simple and **deliberately vague** assertion masks an array of nuanced historical judgments (…).
     c.   (…) government layers charged the appraisals were instead tainted by conflicts of interest and weak or **deliberately inadequate** research.
     d.   The metaphor's two words are deliberately simple and emotive, but challenging.
     e.   The S.E.C. rules are **deliberately conservative** and intended to prevent companies from overstating their reserves.
     f.   They were fake-historical movies, crafted in a **deliberately anachronistic** style to accommodate the stars' gift for eternal-adolescent horseplay.

The Polish equivalent of *deliberately/intentionally* (*celowo*) is also occasionally found with simple adjectives. The examples below have been extracted from the Internet:

(23) a.   **Celowo długie** przerwy pomiędzy aktami skłaniały widzów do rozmowy o ideach spektaklu i przyszłości kraju.
          '**Intentionally long** breaks between acts encouraged viewers to talk about the ideas of the show and the future of the country.'
     b.   Niektóre miejsca mogą być **celowo puste** z powodów „operacyjnych".
          'Some seats may be **intentionally empty** for "operational" reasons.'
     c.   Dla wielu kandydatów jest to trudny do osiągnięcia próg. Jest on **celowo wysoki**, aby zredukować liczbę kandydatów (…).
          'For many candidates, this is a difficult threshold to reach. It is **deliberately high** to reduce the number of candidates (...).'

The apparent ability of Polish *-ny/-czy/-ły* compounds and English adjectival *-ing* compounds to license *deliberately* cannot be taken as indicative of the presence of the external argument in their structure as in the cases where they are preceded with *deliberately*, the agent-oriented modifier is controlled by the copular verb, rather than the compound itself, which is further confirmed by the coordination of *deliberately* with pure adjectives devoid of any verbal element capable of controlling the modifier.

Adjectival synthetic compounds in English and Polish, as shown in the examples below, are also sometimes found to be accompanied by purpose clauses, which is another feature used to diagnose the presence of the external argument (see Sichel 2010; Alexiadou *et al.* 2015).

(24) a.   **These** statements are **thought provoking in order to** make students think deeply about the primary issues in the study of science.
     b.   The police should only be required to police a small group who are deviant and the rest of the citizens must be **law-abiding in order to** ensure a safe society.

(25)   Pokarmy powinny **być lekkostrawne po to**, aby organizm nie marnował zbyt wiele energii na trawienie.
       'Foods should be easily digestible so that the body does not waste too much energy on digestion.'

We find the sentences (24a) and (25) identical to sentences with no traces of the external argument, such as *Grass is green to promote photosynthesis* (Williams 1987; Landau 2000) where

the purpose clause is controlled by a non-agent director or, as is the case with (23b), the purpose clause is controlled by the copular verb. Consequently, we believe that compounds in (24) and (25) are agentless adjectives.

An argument can be made that adjectival compounds project the external argument identical with the described subject; for example, in *thought-provoking book*, the head of the phrase (*book*) can be seen as the initiator of the event of 'provoking' as in the active and passive sentences below: [12]

(26)   a.   John's book provoked the/much thought.
       b.   The/Much thought was provoked by John's book.

The sentences in (25) are possible only in the presence of the additional focus (**Thought was provoked by John's book*) and as such events expressed by them do not fully correspond with -*ing* compounds. Also, a number of -*ing* compounds are characterized by some degree of idiosyncrasy (e.g. *life-giving*); such cannot be turned into passive structures (*life-giving water*; **Life is given by water*).

The relevant linguistic data does not allow us to postulate the presence of the external argument in the structure of Polish -*ny*/-*czy*/-*ły* compounds and English compounds of the *time-consuming* type. This, coupled with the fact that they exhibit a range of non-verbal properties, leads to the conclusion that the examined adjectival compounds are constructed by attaching the adjectivizing suffix directly to the root via RootP which houses the left-most element, which effectively functions as the argument of the root:

(27)   a.   *czasochłonny*   b.   *life-giving*


The syntactic stativity of adjectival compounds effectively means that they express properties rather than events. Given the lack of vP, adjectivizing suffixes -*ny*/-*czy*/-*ły* and -*ing* function as the spell-out of the active semantics of adjectival compounds.[13]

The proposed internal structural is based on Embick's (2003) representation of stative participles which in his framework are identical to simple adjectives:

---

[12]   But note that some compounds, especially those incorpating adverbs (e.g. *quick-acting*), can be used as eventive active participles (*Thanks to the quick acting John, the man was rescued from the fire*).

[13]   The -*ny* suffix can also be the carrier of passive semantics as in the case of *lekkostrawny*.

(28)　empty (Embick 2003: 148)

```
            ASP
           /    \
        ASP      √EMPTY
         |
         Ø
```
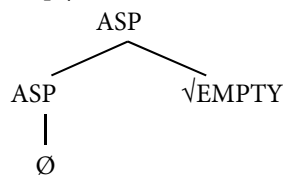
The structural simplicity of *-ny/-czy/-ły* compounds in Polish and adjectival *-ing* compounds in English stems also from the fact that they can only be derived from a limited range of roots and thus are not fully productive. Rather, they appear to be analogical formations. For example, the compound *życiodajny* is the only commonly attested compound in Polish containing the element *dajny* and *chłonny* appears frequently only in *czasochłonny* i *pracochłonny*; other compounds based on these heads are less common.[14] The same holds true for adjectival compounds in English: for example, out of 61 compounds based on the participle *threatening* which appear in the COCA, only 5 occur more than 10 times.[15] Likewise, the COCA returns 36 compounds with *provoking* as the head but only *thought-provoking* is widely employed with 1210 occurrences[16] (the only other compound containing the head *provoking* which appears more than 5 times is *anxiety-provoking* with 71 occurrences). The limited productivity of Polish *-ny/-czy/-ły* compounds and English *-ing* compounds is in line with the assumption put forward by Marantz (2001, 2007 and see also Arad 2003, 2005 and Embick 2010), that unproductive complex words are formed by root attachment.

4. Conclusion

In this paper, we have argued that *-ny/-czy/-ły* compounds in Polish, although appearing to be deverbal, are in fact fully adjectival and non-eventive constructions. Their syntactic stativity is manifested by a range of features typical of adjectives such as the ability to occur in the predicative position, possible coordination with other adjectives, the prefixation with a negating affix and adjectival degree modification. Synthetic *-ny/-czy/-ły* and *-ing* compounds have been shown to be non-agentive, despite the existence of contexts where they are accompanied by the

---

[14]　As pointed out by a *LingBaW* reviewer, *dajny* is also found in compounds such as *złotodajny* 'gold-giving' or *miarodajny* 'lit. measure-giving, reliable'; in a similar fashion, *chłonny* functions as right-most constituent in *energochłonny* 'energy-consuming', *surowcochłonny* 'resource-consuming' or *kosztochłonny* 'cost-generating'. Such compounds are much less frequent than *życiodajny* or *czasochłonny*, with the exception of *miarodajny*, whose meaning is not-transparent. Consequently, an argument can be made that such compounds are analogical formations (analogy has been understood in contrast to productivity, see e.g. Motsch 1981).

[15]　The COCA data are based on the compounds spelled with a hyphen.

[16]　Note that some active participles can give rise to both 'purely' adjectival and deverbal compounds, with only the latter being productive. The participle *giving* functions as the right-most constituent in numerous *-ing* compounds but only *life-giving*, *health-giving* and *pleasure-giving* are commonly attested adjectival compounds. The great majority of compounds based on *giving* belong to the eventive class, e.g. *gift-giving*, *advice-giving*, *order-giving*, *party-giving*, *direction-giving*, *information-giving*, etc. Eventive *-ing* compounds can also be derived from a much wider range of roots. For example, the root √FORM is not found in vp-less adjectival *-ing* compounds but commonly gives rise to deverbal *-ing* compounds (*smog-forming*, *star-forming*, *bone-forming*, etc.).

adverb deliberately and purpose clauses which in such contexts are not controlled by the compound. Polish *-ny/-czy/-ły* compounds have been shown to correspond to adjectival *-ing* compounds in English such as *life-giving*, *though-provoking* or *time-consuming* whose syntactic features are identical with their Polish counterparts.

The stative external syntax of Polish and English adjectival synthetic compounds, further intensified by their limited productivity, leads to the conclusion that their internal structure lacks the vP and VoiceP projections with the adjectivizing suffix merged directly with the lexical root.

References

Alexiadou, A., and E. Anagnostopoulou. 2008. Structuring participles. In C. B. Chang, and H. J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 33–41. Sommerville, MA: Cascadilla Proceedings Project.

Alexiadou, A. and F. Schäfer. 2010. On the syntax of episodic vs. dispositional *-er* nominals. In A. Alexiadou and M. Rathert (eds.), *The syntax of nominalizations across languages and frameworks*, 9–38. Berlin: Mouton de Gruyter.

Alexiadou, A., B. Gehrke, and F. Schäfer. 2014. The argument structure of adjectival participles revisited. *Lingua* 149: 118–138.

Alexiadou, A., Anagnostopoulou E., and F. Schäfer. 2015. *External arguments in transitivity alternations: A layering approach*. Oxford: Oxford University Press.

Anagnostopoulou, E. 2003. Participles and voice. In A. Alexiadou, M. Rathert, and A. von Stechow (eds.), *Perfect explorations*, 1–36. Berlin: Mouton de Gruyter.

Anagnostopoulou, E., and Y. Samioti. 2011. Idiomatic meaning and the structure of participles. In Eliza Kitis, Nikolaos Lavidas, Nina Topintzi, and Tasos Tsangalidis (eds.), *Selected papers from the 19th International Symposium on Theoretical and Applied Linguistics*, 101–110. Thessaloniki: Monochromia.

Arad, M. 2003. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory* 21: 737–778.

Arad, M. 2005. *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Springer.

Bruening, B.. 2014. Word-formation is syntactic: Adjectival passives in English. *Natural Language and Linguistic Theory* 32(2): 363–422.

Cetnarowska, B. 2017. Group adjectives, argument structure and aspectual characteristics of derived nominals in Polish and English. In M. Bloch-Trojnar, and A. Malicka-Kleparska (eds.), *Aspect and valency in nominals*, 131–155. Berlin & Boston: Walter de Gruyter.

Davies, M., 2008. *The Corpus of Contemporary American English (COCA): 560 million words, 1990–present*. Available online at https://www.english-corpora.org/coca/.

Embick, D. 2003. Locality, listedness and morphological identity. *Studia Linguistica* 57(3): 143–169.

Embick, D. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3): 355–392.

Embick, D. 2010. *Localism vs. globalism in morphology and phonology*. Cambridge, MA: The MIT Press.

Gehrke, B. 2011. Stative passives and event kinds. In I. Reich, E. Horch, and D. Pauly (eds.), *Sinn und Bedeutung 15: Proceedings of the 2010 Annual Conference of the Gesellschaft für Semantik*, 241–257. Universaar – Saarland University Press: Saarbrücken.

Kennedy, C., and L. McNally. 1999. From event structure to scale structure: Degree modification in deverbal adjectives. In T. Matthews, and D. Strolovitch (eds.), *Proceedings from Semantics and Linguistics Theory IX*, 163–180. Ithaca, NY: CLC Publications.

Kratzer, A. 2000. Building statives. *Proceedings of the Berkeley Linguistics Society* 26: 385–99.

Landau, I. 2000. *Elements of Control*. Dordrecht: Kluwer.

Levin, B., and M. Rappaport. 1986. The formation of adjectival passives. *Linguistic Inquiry* 17(4): 623–661.

Marantz, Alec. 2001. Words and things. Handout, MIT.

Marantz, A. *2007*. Phases and words. In S. H. Choe (ed.), *Phases in the theory of grammar*, 191–222. Seoul: Dong in Publisher.

Meltzer-Asscher, A. 2011. *Adjectives and argument structure*. Ph.D. diss. Tel Aviv University.

McIntyre, A. 2013. Adjectival passives and adjectival participles in English. In A. Alexiadou, and F. Schäfer (eds.), *Non-canonical passives*, 21–42. Amsterdam: John Benjamins.

Motsch, W. 1981. Der kreative Aspekt in der Wortbildung. In L. Lipka (ed.), *Wortbildung*, 94–118. Darmstadt: Wissenschaftliche Buchgesellschaft.

Pęzik, P. 2012. Wyszukiwarka PELCRA dla danych NKJP. A.Przepiórkowski, M. Bańko, R. L. Górski, and B. Lewandowska-Tomaszczyk (eds.), *Narodowy Korpus Języka Polskiego*. Wydawnictwo PWN.

Sichel, I. 2010. Event structure constraints in nominalization. In A. Alexiadou, and M. Rathert (eds.), *The syntax of nominalizations across languages and frameworks*, 151–190. Berlin & New York: Mouton de Gruyter.

Szymanek, B. 2010. *A panorama of Polish word formation*. Lublin: Wydawnictwo KUL.

Wasow, T. 1977. Transformations and the lexicon. In P. Culicover, A. Akmajian, and T. Wasow (eds.), *Formal Syntax*, 327–360. New York: Academic Press.

Williams, E. 1987. Implicit arguments, the binding theory, and control. *Natural Language and Linguistic Theory* 5: 151–180.