LingBaW

A peer-reviewed online journal

czasopisma.kul.pl/LingBaW

LingBaW

Linguistics Beyond and Within

VOLUME 8 (2022)

THE JOHN PAUL II CATHOLIC UNIVERSITY OF LUBLIN KUL 1918

TYPESETTING: Karolina Drabikowska

COVER DESIGN: Wojciech Malec

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

e-ISSN: 2450-5188

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

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

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

Articles

LingBaW *Linguistics Beyond And Within 8 (2022), 5-23*

Phenomena in Romance verb paradigms: Syncretism, order of inflectional morphemes and thematic vowel*

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

Abstract
This article aims to propose a treatment of the internal morphological organization of words, based on the idea that morphology is part of syntactic computation. We disagree with Distributed Morphology model, whereby morphology is identified with a post-syntactic component conveying an information 'separated from the original locus of that information in the phrase marker' (Embick and Noyer 2001: 557) by rules manipulating syntactic nodes. We also consider inadequate the costly and complex syntactic structures that cartographic approach maps into inflectional strings. We pursue a different conceptualization assuming that morphology is governed by the same rules and principles of syntax. Sub-word elements, including inflections, thematic exponents and clitics, are fully interpretable and enter (pair-)merge operations (in the sense of Chomsky 2020a,b, 2021) according to their content, giving rise to complex words.

Keywords: morphology, syntax, inflection, thematic vowel, Romance languages

1. Is there a morphological component?

In the generative framework, the best-known generalization concerning the distribution of inflectional morphemes is Baker's (1988) Mirror Principle, whereby the verb moves to combine with the closest suffix: V attaches to T, and then T-V moves to AgrS, which closes the complex word. Thus, for instance, the 2nd plural of the Italian imperfect *lava-va-te* 'you(pl) washed', is obtained by moving *lava-*, the lexical head to the inflectional head T/I, and then to the agreement position for the subject, where *–te* is taken on. Therefore, the complete form is the result of syntactic derivation. The Mirror Principle translates into syntactic operations the idea, traditional in generative grammar, that the composition of complex words is an ordered

* This contribution is the result of a common work, and specifically: conceptualization, B. Baldi and L.M. Savoia; methodology, B. Baldi and L.M. Savoia; data curation, L.M. Savoia; writing—original draft preparation, B. Baldi; writing—review and editing, L.M Savoia. The data we discuss in this article were collected through field investigations with native speakers by the authors or, when indicated, they come from Manzini and Savoia (2005).

We are grateful to our informants for their patient and intelligent cooperation.

cyclic mechanism. At once, it realizes the linkage between inflection and syntactic categories. Differently, Distribute Morphology (DM, see Halle and Marantz 1993, 1994), the more adopted approach to morphology, identifies morphology with an autonomous component, in which the insertion of morphemes is after syntax (Late insertion) and yet it is based on a mechanism in which sub-word elements (affixes and clitics), are understood as 'dissociated morphemes' conveying an information 'separated from the original locus of that information in the phrase marker' (Embick and Noyer 2001: 557) and involving post-syntactic rules of the linear adjacency (Local dislocation) (Embick and Noyer 2001). As we can expect, there are empty morphological elements, as in the case of Thematic Vowels of Romance languages, identified with 'ornamental pieces of morphology' by Embick (2010). Moreover, agreement and case morphemes are not represented in syntax, but they are added postsyntactically 'during Morphology'. As for 'Late insertion' Collins and Kayne (2020) underline that, apart from other critical points, it entails that the relevant features are repeated in Vocabulary Items and as terminals of syntactic representations; moreover, the syntactic features are duplicated in post-syntactic insertion of phonological forms.

Let us deepen the role of Thematic Vowels (TV) in the verbal inflection of Romance varieties. Usually, tense/ aspect/ mood exponent(s) are inserted in-between the root, possibly enlarged by the TV, and person and number inflection, as in (1) for Italian. The R underwritten indicates the root.

(1) *dorm- i va- te*
sleep$_{R}$ TV Tense AgrS$_{2PL}$
'you (pl) slept'

In DM treatment, *-te* is the morpheme introduced for the feature 2pl, *-va-* for T$_{Impf}$ whereas *-i-* is introduced by an ad hoc rule whose effect is to specify the class of the verb, a post-syntactic semantically empty mechanism as (2a). The motivation for rules like (2a) is to create the correct context for the subsequent Vocabulary insertion, (2b).

(2)
a. $\varnothing \rightarrow i_{TV}$ / Root$_{4thClass}$ __
b. Impf → -v- / TV __

The question is why language, understood also in the broad sense (Hauser, Chomsky, and Fitch 2002), should be engaged in obscuring or complexing the relationship between interpretive (IC) and sensory-motor (SM) interfaces (Manzini and Savoia 2011a, 2018). The idea that morphology is someway an imperfection of language is rooted in morphosyntactic literature. It is no accident if Aronoff (1998: 406) concludes that 'morphology is inherently unnatural, it is a disease, a pathology of language'. But, as we shall see, we do not believe that. On the contrary, it is possible to think of the relation between syntax and morphology as based on the lexical content of those 'pieces'.

The approach to morphology that we will follow, is based on the idea that morphology is part of the syntactic computation and there is no specialized component for the morphological structure of words (Manzini and Savoia 2017, 2011a, Manzini et al. 2020,

Savoia et al. 2018; see also Collins and Kayne 2020). Lexical elements, including morphemes, are fully interpretable, and contribute to externalizing the syntactic structure. Hence, morphemes are endowed with interpretive content and obey the general requirement of *inclusiveness condition* on syntax, thus excluding Late Insertion and the other adjustments provided by Distributed Morphology, such as the manipulation of terminal nodes, impoverishment, and fusion of φ-features.

Inflected words are analyzed as the result of the Merge operation that combines inflectional heads with a category-less lexical root R, interpreted as a predicate. In the case of nominal elements, inflectional contents are Class (gender feminine/masculine) and other classificatory properties such as number and case (Manzini and Savoia 2011b). In inflected verbal forms, agreement features and mood/ tense/ voice inflections are merged with R. Specifically, syncretism and other kinds of ambiguity imply a treatment based on the interpretive properties of the items/inflectional exponents and not on different syntactic structures. We distance ourselves also from the attitude of cartography to build the meaning through and in the terms of the structure. Similar conclusions are now further supported by Wood and Marantz (2011), and specifically in Collins and Kayne (2020) with regard to the relationship morphology/ syntax.

We can assume that the Merge operation (Chomsky 2020a,b) in (3) creates the combination of morphemes in complex words:

(3) Merge(X,Y) → [X,Y]

This procedure operates on syntactic objects, therefore accessible both to semantic interpretation and phonological linearization (see also the discussion in Collins and Kayne 2020: 13). Chomsky (2020a: 55) sees in pair-merge the way of treating head raising: 'It's always described incorrectly. If a verb raises to inflection, say to T, it's always described as if the T-V complex becomes a T; but it's not, it's a V-the outcome of the adjunction is really verbal, not inflectional'. As for modification as in the case of adnominal adjective expressions such as *young man*, Chomsky concludes that it is the result of an operation of conjunction where the same categorizer *n* (Link) is shared by the conjuncts, whereby R(oot) merges with the Link/categorizer *n*. In this line, we conceptualize categorizers such as *v, n*, as the bundles of φ-features that characterize the functional content of words entering into the agreement operations (Manzini 2021, Baldi and Savoia 2021). Hence, *n* is the label for the class and number features of the nominal agreement, and *v* is the label for the verbal categories of tense, aspect, and mood that make an eventive/stative root a verb. As it is well known also from a typological perspective, the inflection, for instance of tense or agreement, is sufficient to make a root a verb from a noun or, conversely, a noun from a verb.

Chomsky (2021: 30 and 36 ff.) in discussing the illegitimate nature of head movement, observes that V-to-T raising is unjustified because 'interpretation is the same whether a verb raises to INFL or stays *in-situ*'. He proposes that head movement must be treated as the result of Merge between root and inflection, on which an operation of amalgamation applies, yielding complex forms subject to externalization. Now, the external argument is interpreted in the phase of T by the inflected form of the verb, and v is no longer involved in the process.

Morphology substantially involves the combination of morphemic heads with roots or morphological amalgams, also merging clitics to words. In the minimalist model we apply, the agreement is accounted for as the morphological manifestation of the identity between referential feature sets corresponding to the arguments of the sentence. In other words, there is no uninterpretable category triggering the raising of a goal (see Chomsky et al. 2019, Chomsky 2020a,b). If words are formed by combining the uncategorized lexical root with morphological elements, inflectional morphemes select for the compound including the root and its immediately attached morpheme (cf. Marantz 2001, 2007). In (4), a partial contextualization of the occurrence of TV is suggested for the second plural of the present indicative and the imperfect; in (4) φ stands for the relevant features.

(4) $\text{Root}_{\text{4thClass}}$ __ present 2nd plural/v_{Impf} ←→ $i_{\text{TV}/\varphi}$

This model, therefore, excludes the separation between inflectional morphology, introduced in syntax, and derivational morphology, substantially lexical as generally in the generative approach. The idea is that 'syntax perform[s] all merger operations including those between morphemes within a word' (Marantz 2001: 6).

In the following sections, we aim to show that identifying morphology with syntactic Merger gives rise to valuable results. To do this, we will present and discuss some crucial morphological phenomena concerning the verb in Romance varieties: syncretism and the interaction between thematic vowel and inflection.

2. Syncretism

As widely known in the relevant literature, in Romance dialects spoken in North Italy the subject is expressed by a clitic pronoun (SCl) preceding the inflected verb. SCls can cover the entire paradigm or occur only on a subset of persons, as illustrated in Manzini and Savoia (2005). Interestingly, Renzi and Vanelli (1983) suggest some (weak) implications, whereby if there is only one SCl, this is the SCl of 2ps, and if there are two SCls, they are the 2ps and 3ps, although a sufficiently large sample shows that mandatory constraints do not exist (Manzini and Savoia 2005).

Several syncretism phenomena appear that can also involve both the inflection and the SCls paradigm. We will examine a case in point in order to explore the properties of syncretism and its theoretical treatment, addressing the distribution of inflections and subject clitics in the dialect of Trecate, spoken in the area between Piedmont and Lombardy. The examples in (5a,b,c,d) illustrate the four verb classes. As we see, the SCls coincide in form *i* except for the 2nd/3rd persons, in (5e), and the inflection *–a* covers four persons, as schematized in (5f). We remain in doubt about the inflection of 2nd plural, given that in the present indicative this is the only form including the realization of the thematic vowel, with the result that it is the latter that specifies the person rather than a specialized inflectional exponent. We will come back in the following section to syncretism in the thematic part of the 1st and 2nd plural persons, where *-u-m* is generalized to all classes, while *–i* of the 2nd/3rd/4th classes contrasts with *-ɛ* in the 1st class. In (5a',b',c',d') the infinitive is provided.

(5)

a.	b.	c.	d.
i tʃam-a	i vød-a	i pɛrd-a	i drøm-a
te tʃam-a	te vød-a	te pɛrd-a	te drøm-a
a tʃam-a	a vød-a	a pɛrd-a	a drøm-a
i tʃam-um-a	i vid-um-a	i pird-um-a	i drum-um-a
i tʃaˈm-ɛ	i viˈd-i	i pirˈd-i	i druˈm-i
i tʃam-u	i vød-u	i pɛrd-u	i drøm-u
'I call, etc.'	'I see, etc'	'I loose, etc.'	'I sleep, etc.'
a' tʃaˈm-ɛ 'to call'	b'. vøt 'to see'	c'. pɛrd 'to lose'	d'. druˈm-i 'to sleep'

e. SCls

i	←→	1st sg / 1st pl / 2nd pl / 3rd pl
te	←→	2nd sg
a	←→	3rd sg

f. Inflections

-a	←→	sg PresInd/ 1st pl
-u	←→	3rd pl
-um-	←→	1st pl
-i[+stress]	←→	2nd pl

Trecate

For the sake of comparison, we report the data of other Gallo-Italic varieties ((6a,b) Southern Piedmont and (6c) Lombardy) concerning the present indicative of 'sleep'. We note that in some dialects the 3rd person SCl distinguishes a masculine and a feminine form: we find the masculine *u* and the feminine *a* in (6a) and (6b), and the masculine *al* and the feminine *l-a* in (6c). In (6c) two plural SCls are also attested, i.e., the masculine *i* and the feminine *l-i* (cf. Manzini and Savoia 2005).

(6)

a.	b.	c.
a drø:mu	a drɔ:m	a dɔrmi
ti drø:mi	ta drɔmi	at *d*ɔrmi
u/a drø:ma	u/a drɔmæ	al/la *d*ɔr*ʊ*m
a drumimu	a ˈdrɔmʊm	a dʊrmem(*a*)
a druˈmi	a druˈmi	a dʊr ˈmi
i drøma	a drɔmen	i/li dɔr*ʊ*m
'I sleep, etc.'		
Arquata Scrivia	Viguzzolo	Revere

(6) shows two general properties of paradigms: (i) syncretism in SCls is widespread, independently of the properties of the inflectional paradigm; (ii) an implicational relation between the specialized SCls and inflections seems to concern the fact that nondistinct SCls generally combine with specialized inflections. Inversely, we can find distinct inflections combining with distinct SCls, as in the case of the 2nd singular person, mostly characterized by a specialized SCl including the morpheme *t*, and by a specialized inflection *–i*.

Let's move on to how syncretism can be conceived and treated. Syncretism is a typical test bench in morphological analysis, generally seen as supporting the hypothesis of a formal level independent of semantics and syntax, the 'morphomic' level of Aronoff (1994). DM deals with syncretism in terms of 'Late Insertion' of the Vocabulary items and manipulation of the features by impoverishment rule, with the consequence that morphemes/ lexical entries are inserted after modifying the underlying syntactic terminal nodes. Thus, for instance, the

occurrence of the ending *–a* in (5f) can be viewed as the insertion of a default empty element, in (7b), in impoverished contexts like (7a).

(7)
a. sg/ 1st pl → ∅ / [Pres Ind] __
b. ∅ → a

Collins and Kayne (2020) propose an analysis of syncretism based on silent elements not externalized, while the morpheme realized is the same. So, in the case of the syncretic masculine and feminine reading of English *they*, the rule in (8a,b) is formulated (Collins and Kayne 2020: 16), where *–ey* is analysed as an irregular form corresponding to two gender contexts, phonologically unexpressed.

(8)
a. [DP th- [#P -ey [GenP MASC NP]]]
b. [DP th- [#P -ey [GenP FEM NP]]]

The silent contexts are obscured by the morpho-phonological idiosyncrasies of the pronoun.

In other words, morphology has recourse to markedness criteria whereby a very spread unspecialized element is treated as the default. As we know, markedness is a circular notion, so what is frequent is unmarked and what is unmarked is frequent. Moreover, in many cases, it is undecidable, as in the case of the pronoun of 1st person, marked with respect to the unspecialized nature of the 3rd person, but clearly unmarked with respect to the discourse. It is no accident if it gives rise to widespread Differential Marking phenomena both of subject and object. In the case in point, default elements are substantially unmotivated insofar as they reflect a null content. The question of why languages should obscure the linkage between morpho-syntax and interpretation remains unexplained.

2.1. The interpretative properties of morphemes

We may think that the suitability of certain forms to operate in correspondence with different interpretive properties derives from the semantic content of such forms. Morpho-syntactic contexts can contribute to introducing different readings (cf. Franco et al. 2020: 12). As to (5a), we initially have to establish the status of SCls. A reasonable idea is that clitics are an expression of the agreement properties of verbal projections, as also recently argued by Roberts (2010, 2018). According to Roberts (2018: 261), SCls are the result of a sort of fission (in the sense of DM) of φ-features of T, which in languages with obligatory SCls, like Northern-Italian dialects, gives rise to the distribution in (9).

(9) SCl_{φ} [T + φ]
Fission: {φ} in the context [T__, is changed to φ [T+ φ]

The gist is that SCls are a morphological tool doubling φ-features agreement on the verb in T, but as a sort of reflex of the agreement (Baldi and Savoia 2021). If that was exactly the case, we would expect an interpretive bi-univocal correspondence between SCls and inflections. The

examples in (5) and (6) show that things are different. Moreover, Chomsky (2015: 9, 2020a) has clarified the fact that the rich agreement of the verb is sufficient to label < φ, φ> both TP and {Spec TP}, hence refusing the traditional approaches based on the little *pro* and the connected treatments of SCls. However, the agreement between SCls and inflections is a necessary requirement, given that SCls and inflections realize the same argument. Doubling of features does not implies a perfect formal correspondence between SCls and inflections, but it is based on compatibility with the same referent, as observed and discussed, for instance for the relation between D and the agreement inflection on N in Savoia et. al. (2019), Manzini et al. (2020).

Pair-merge yields well-formed combinations that are amalgamated into the corresponding inflected forms, as in (10a). Merging Infl and R creates the inflected verb, whose inflection identifies the subject. (10b) creates the sequence *SCl+V* on the basis of sharing φ-features.

(10)
a. $< \text{Infl}_{\varphi}, \text{R} > \rightarrow [_{\text{T}}\text{R} + \text{Infl}_{\varphi}]$
b. $< [_{\varphi}\text{SCl}], [_{\text{T}}\text{V} + \text{Infl}_{\varphi}] > \rightarrow [_{\text{T}} [_{\varphi}\text{SCl}] [_{\text{T}}\text{R} + \text{Infl}_{\varphi}]]$

But what are the features involved? Syncretism seems to indicate some type of underspecification, so to say, and it must also be such that SCLs and inflectional elements are combined as markers of the same argument. As to the SCl *i*, we recall that in the Trecate dialect, as well as in the other Northern Italian dialects, it is also the OCl plural and the dative, as illustrated in (11a,b)

(11)
a. l a tʃama-r/ ra/ i
SCl has called-him/ her/ them
'(s)he has called him/ her/ them'
b. da-i kust
give-him/her this
'give it to her/ him'

According to a proposal by Manzini and Savoia (2011a, 2017) and following a suggestion by Chierchia (1998), plural morphology can be associated with the part-whole/sub-set content, i.e., the inclusion relation [⊆], as in (12). In other words, the content of the plural, [⊆], indicates that the argument of the root, here represented by the SCl, can be partitioned into subsets. This analysis allows explaining the coincidence between plural and dative in many Romance varieties, as in (11). More precisely, in the case of the plural, the part-whole relation concerns the argument of the nominal/ pronominal base, while in the case of dative, the sub-set relation is between the possessum and the beneficiary of the event. Therefore, dative can be treated as a type of 'zonal inclusion' in the sense of Belvin and den Dikken (1997)[1]. In fact, while the nominal plural expresses a subset within the set of the individuals

[1] The idea is that in all possession contexts the conceptual property of 'inclusion' is involved, in the sense initially discussed in Manzini and Savoia (2011a) whereby all types of possession fall under the same basic relation. Their proposal resumes the idea of Belvin and den Dikken (1997:170) according to whom 'entities

identified by the noun, the dative expresses the inclusion relation between the object of the verb and the recipient, i.e., the dative.

(12) i = [⊆]

Syncretism in the persons implies that *i*, i.e. the exponent for the sub-set relation [⊆], possibly includes the singular in the case of the first person here, where the 1st person is evidently treated as a subset of the participants.

On the contrary, the syncretic inflection *–a* in the paradigm in (5) can be characterized as an element of definiteness, which selects all the persons except the second and the third plural, as suggested in (13). In this dialect, the 3rd plural has the specialized inflection *-u*; as to 2nd plural, we will see that it can be identified with the TV.

(13) a = Def, 1,2,3,4,6 person __

It is interesting to note that *-a* is also merged with the specialized exponent of the first plural, as in (14a,b), where the amalgam *drum-um-a* 'we sleep' is yielded by Merge and the process of amalgamation, that we directly represent in the derivation.

(14)
a. $< [drum_R], um_{1pl} \rightarrow [_{\varphi} drum+um_{1pl}]$
b. $< [_{\varphi} drum+um_{1pl}], a_{Def} \rightarrow [_{T/\varphi} [drum+um_{1pl}] a_{Def}]$

As for its interpretation, a complex realization such as *vid-um-a* 'we see' puts together the specification for the inclusion, ⊆, with Def and person specifications, as in (15), exploiting their referential compatibility.

(15) $[_{\subseteq} i] [_T [drum\text{-}um_{1pl}] a_{Def}]$

At this point, the main question is why the 2nd and 3rd persons require a specialized SCl, differently from the other, a question regarding DSM, and, by hypothesis, relating to the 'third factor'[2] and other cognitive constraints.

3. Thematic vowel

Let us consider now the thematic vowel TV, i.e., the vocalic element that in Romance languages is generally associated with the inflectional class (conjugation) of the verb. For

have various zones associated with them, such that an object or eventuality may be included in a zone associated with an entity without being physically contained in that entity [...] The type of zones which may be associated with an entity will vary with the entity'. Hence, possession–on a par with location–can be understood as a type of 'zonal' inclusion.

[2] Chomsky (2005: 6) calls third factor the set of 'Principles not specific to the faculty of language' operating into the growth of language in the individual. Among others, they include 'principles of structural architecture'. We can think that the structural architecture interacts with other cognitive mechanisms of a general nature in the conceptualizing the world.

instance, in Italian *-a-* characterizes the first conjugation, as in *port-a-re* 'to bring', *port-a-v-o* 'I brought', *port-a-t-o* 'brought *past participle, msg*', *-e-* the second and third conjugations and *-i-* the fourth one. The thematic vowel raises reconstructive and interpretive problems, as far as it is related neither to a clear etymological origin[3] nor to some salient semantic status, save to distinguish verbal classes, even if with some important complications. Moreover, in Romance languages the thematic vowel is subject to many distributional idiosyncrasies with phenomena of extension or gaps in paradigms, and in the 2nd and 3rd conjugation is generally excluded in strong perfects and participles, where the inflection of tense or aspect is directly combined with the root, such as in *perd-e-re* 'to lose', but *per-s-i* 'I lost', *per-s-o* 'lost' (cf. Calabrese 2015 for a DM approach to Italian strong paradigms).

3.1. Extension phenomena of TV in Northern Italian dialects

Let us begin with the perhaps best-known process concerning TV, i.e. the extension of a thematic vowel to all verb classes as attested in many Northern and Central Italian dialects, and in other Romance languages. Typically, it is the TV *-e-*, i.e. the thematic vowel characterizing the second and third conjugations which extends to the verbs of first and fourth conjugation. We illustrate a pattern of this type provided by the North Piedmontese dialect of Trecate (cf. (5)) in (16) and the adjacent Lombard dialect of Casorezzo, in (17). (16a,b,c,d) illustrate the distribution of *-e-* in the imperfect indicative for the four verb classes, where the extension of *-e-* is accompanied by the complete syncretism in plural persons.

(16)

a.	b.	c.	d.
i tʃam-e-v-a	i vid-e-v-a	i pird-e-v-a	i drum e-v-a
te tʃam-e-v-a	te vid-e-v-a	te pird-e-v-a	te drum-e-v-a
a tʃam-e-v-a	a vid-e-v-a	a pird-e-v-a	a drum-e-v-a
i tʃam-e-v-u	i vid-e-v-u	i pird-e-v-u	i drum-e-v-u
i tʃam-e-v-u	i vid-e-v-u	i pird-e-v-u	i drum-e-v-u
i tʃam-e-v-u	i vid-e-v-u	i pird-e-v-u	i drum-e-v-u
'I called, etc.'	'I saw, etc.'	'I lost, etc.'	'I slept, etc.'

Trecate

A similar distribution characterizes Casorezzo, as in (17a,b,c,d).

(17) Imperfect indicative

a. a parl-e-(v)-u, te parl-e-(v)-a, al parl-e-(v)-a, a parˈl-e-(v)-am/um, a parl-i-(v)-i, a parˈl-e-(v)-an 'I spoke, etc.'

b. a vid-e-(v)-u, te vid-e-(v)-a, al vid-e-(v)-a, a viˈd-e-(v)-am/um, a vid-i-v-i, a viˈd-i-(v)-an 'I saw, etc.'

c. a rid-e-(v)-u, te rid-e-(v)-a, al rid-e-(v)-a, a riˈd-e-(v)-am/um, a rid-i-v-i, a riˈd-e-(v)-an 'I laughed, etc.'

d. a durm-e-(v)-u, te durm-e-(v)-a, al durm-e-(v)-a, a durˈm-e-(v)-am/um, a durm-i-v-i, a durˈm-e-(v)-an 'I slept, etc.'

Casorezzo

In the present indicative of the dialect of Casorezzo, the second plural coincides across all classes as schematized in (18a), repeating a distribution already seen in the paradigms in (5).

[3] Villanueva Svensson (2021) connects thematic vowels in Indo-European verbs with a denominative function from thematic adjectives. However, the question is very controversial.

By contrast, the first plural present indicative is stressed on the root and lacks the thematic element, as in (18b).

(18) Present indicative
a. a parˈl-i: 'you speak'
a viˈd-i: 'you see' '
a riˈd-i: 'you laugh'
a durˈm-i: 'you sleep'
b. a parl-um/am 'we speak'
a ved-um/am 'we see'
a rid-um/am 'we laugh'
a dɔrm-um/am 'we sleep'

Casorezzo

The comparison between the imperfect indicative of Trecate in (16) and that of Casorezzo in (17) highlights an interesting point, i.e., that in the 2nd plural, illustrated in (18a), the TV *-i-* occurs as the specialized exponent. In other words, TVs can express both tense/ aspect features and/or person features, exactly like inflection. This is clearly evidenced by the data concerning some Lombard-Alpine dialects, such as, for instance, those of Sonogno (Verzasca Valley) and Cavergno (Maggia Valley), spoken in the Switzerland area bordering Northern Italy (data from Manzini and Savoia 2005). Like (16) and (17), the imperfect extends the same TV *-ɛ-* to the first class in Sonogno in (19), differently from Cavergno in (20), where *-a-* is retained. Nevertheless, in both, the 2nd person has the specialized TV *-i-* (highlighted in grey).

(19) a tʃam-ɛ-v-a 'SCl called'/ a ved-ɛ-v-a 'SCl saw'
ti tʃam-i-v-a
o tʃam-ɛ-v-a
o m tʃam-ɛ-v-a
a tʃam-i-v-o
i tʃam-ɛ-v-a

Sonogno

(20) a tʃam-a-v-a 'SCl called'/ a veɟ-ɛ-v-a 'SCl saw'
ti tʃam-i-v-u
u tʃam-a-v-a
u m tʃam-a-v-a
a tʃam-i-v-u
i tʃam-a-v-a

Cavergno

A similar distribution is attested in the present indicative in (5) and (18a), where the TV *-i* is specialized for the 2nd plural in all classes in (18a) and in all but first class in (5). The comparison between the first plural present indicative in (18b) and in (5) shows that the same string *–um* can be treated as a usual inflectional element.

(21)
a. drum-ù-m-a 'we sleep' Trecate
b. dɔrm-u-m 'we sleep' Casorezzo

The data synthesized in the comparison between (21a) and (21b) highlights the possibility for the thematic morphemes to behave like nothing but a type of inflection. In *dɔrm-um* 'we sleep' in (21b), the rhizotonic nature of the form suggests that the morpheme *–u-m* is treated as a simple inflectional ending. However, it includes a clear inflectional part, that is *–m*, and an element *–u-* reminiscent of the thematic vowel, otherwise attested in many Piedmontese dialects, as in *drum-ù-m-a* 'we sleep' in (21a). We conclude that in (21b) an inflectional exponent replaces the thematic vowel, so confirming the close structural proximity between inflection and TV.

3.2. TV as a nominal property

As discussed in section 1, in DM thematic vowels have been seen as empty elements (Embick 2010), a conclusion made own by Calabrese (2015) where thematic vowels are 'special morphological elements adjoined to certain functional heads in morphological structure' by means of an ad hoc rule. In fact, this analysis is common to different formal approaches to morphology, at least starting from Anderson (1992) and Aronoff (1994), where theme vowels are considered 'empty morphs':

> The theme vowel is thus a marker of the category *verb* only in the sense that it is determined by the category *verb*, […] In itself, it has no significance. It is empty. Nonetheless, it is not useless. It has a use in the language, but that use is purely morphophonological: the theme vowel is the *conjugation vowel*, it serves to determine the conjugation of the verb stem, or which inflectional affixes will realize the various morphosyntactic properties that the verb bears in a particular instance.
>
> Aronoff (1994: 44)

Maiden (2018), using the notion of 'morphome' proposed in Aronoff (1994) as the autonomous representation level of the morphological elements, in turn defines the thematic vowel as 'a referentially empty element present in some cells of the in*fl*ectional paradigm and is the basis of the traditional distinctions by in*fl*exion class.' As Thornton (2005) underlines, assuming empty morphs weakens the adequacy of models where morphologically complex forms are seen as composed by the concatenation of pieces endowed with semantic content. Moreover, it remains true that empty elements make the morphological facts substantially unexplainable or irrelevant for linguistic analysis, except for vague functional criteria.

In what follows, we see that TVs are able to behave like the other functional exponents. More precisely, TVs can extend over different or all classes in correspondence to interpretive properties, not only tense and/ or mood, but also number and person, and, in participles they can host the nominal agreement. We note that all these behaviors are manifested also by inflectional exponents, such as person and number, while in participle the nominal agreement has the usual role of inflectional exponents. Moreover, it is usually the case that inflection is extended through all classes, such as *-o, -i* for the 1st and the 2nd person of present indicative. Nevertheless, specialized inflections are also able to signalize tense properties, as in the perfect in Standard Italian, where we find personal specialized exponents, as in (22).

(22) port-a-i, port-a-sti, porˈt-ɔ, port-a-mmo, port-a-ste, port-a-rono 'I brought, etc.'

In (22), specialized inflections for the perfect are combined with thematic vowels, i.e. *-i-* in the 1st sg, *-sti* in the 2nd sg, *-mmo*, *-ste* and *–rono* in the three persons of the plural. In addition, in the 3rd sg the theme and the person overlap in a unique form. In the next section, we will explore a similar distribution in a Central Italian dialect.

Before illustrating the phenomena concerning TV, it must be noted that we consider realizations of the thematic vowel only the stressed vocalic elements interpolated between the root and the tense/mood/aspect or the person/number exponent, i.e., the thematic vowels recognizable as pieces of the morphological string. An interesting topic is why *-e-* is favored in the mechanisms of extension, as shown by the dialects examined below as well as in the imperfect indicative of French (for a historical point of view, cf. Nyrop 1903). A plausible hypothesis is that the TV *-e-* actually was already specialized in the original 2nd and 3rd conjugations, given that in those classes perfects and participles excluded *-e-*, since they were strong or formed by the TV *-u-*. The distribution of -e- was concentrated in the imperfect, independently from the infinitive where all original classes were represented. Again, causes concerning the cognitive implementation of verbal properties seem to explain why the imperfect and other past or Irrealis forms are more subject to morphological reorganization processes implying the manipulation of TVs and inflections. We can link the greater resistance of the present indicative with its Discourse Linked nature, deictically working in establishing the interpretive properties of the event. The result seems to be that the basic distinctions, including the contrast between inflections and TVs, are preserved.

Resuming and deepening a proposal of Manzini and Savoia (2005, 2007, 2011a) we can identify TV with an N element, i.e., a sort of nominalizer that changes R into an inflected base, substantially like the other agreement inflections of the verb. More precisely, the hypothesis adopted here is that Thematic Vowels introduce an indefinite variable 'x', whose value is fixed by the subject. In other words, TVs are nominal inflections making the verbal root a predicate available to combine with the tense/aspectual/modal elements.

Let us take a form such as the 2nd plural of the imperfect indicative in (17a) *ʧam-e-v-u* 'you(pl) called'. It is the result of merging operations that combine TV and R, with the effect of assigning to R a nominal slot x, as in (23a). This amalgam is merged to the tense/ aspect exponent *-v-*, yielding (23b). The agreement inflection realizes the φ-features of T, in (23c).

(23)

a. $< ʧam_R, e_x > \rightarrow [[ʧam_R]\ e_x]$

b. $[_v < [[ʧam_R]\ e_x, v_{PastProgressive}] > \ldots \rightarrow [_v\ [[ʧam_R]\ e_x]\ v_{PastProgressive}]\ \ldots$

c. $[T < [[ʧam_R]\ e_x]\ v_{PastProgressive}], i_{Infl} > \rightarrow [_T\ [[[ʧam_R]\ e_x]\ v_{PastProgressive}]\ i_\varphi] \ldots$

-e- is associated with tense/ aspectual proprieties, as suggested by the selection restriction in (24), whereby the derivation is good only if (24) is fulfilled.

(24) $e_{TV} \leftrightarrow \_\_\ v_{PastProgressive}$

We can understand a constraint of the type in (24) as a property learned by the speaker. Its application relies on the 'free application' of Merge in the sense discussed in Chomsky (2020b) and Chomsky et al. (2020). The case in which the TV includes also a specialized cluster of φ-

features, as the 2nd person features introduced by the morpheme *-i-* in (18), (19a), (20), and (21), for instance, *parl-i-(v)-i* 'you spoke', is more restrictive. We conclude that the only possible occurrence of *-i-* of 2nd person is when the subject is of 2nd person in turn, as in (25a-c).

(25)

a. $< \text{parl}_R, i_{2ps} > \rightarrow [_{2ps} [_R \text{parl}] i]$

b. $[_v < [_{2ps} [\text{parl}_R] i], v_{PastProgressive} > \ldots \rightarrow [_v [[\text{parl}_R] i_{ps}] v_{PastProgressive}] \ldots$

c. $[_T < [[\text{parl}_R] i_{2ps}] v_{PastProgressive}], i_{2ps} > \rightarrow [_T [[[\text{parl}_R] i_x] v_{PastProgressive}] i_{2ps}] \ldots$

In (25) the feature 2ps of the inflection *-i* agrees with the feature 2ps of TV. The ability of the TV *-i-* to express 2ps agreement features characterizes also the 2nd plural person of the present indicative in (5), except for the first conjugation, and in (19). Similar conclusions have been reached in the discussion around (21), in the sense that inflectional exponents and thematic vowels share similar referential properties.

3.3. TVs and agreement inflections

All in all, what we see is that in many varieties TVs do not provide the subdivision of verbs into purely formal classes, but systematically introduce interpretive properties concerning persons instead of/ in addition to tense/ aspect distinctions. In other words, TVs are not simple signals of set of roots (conjugations) but contribute to semantics on a par with the inflectional suffixes. The verb pattern of the Abruzzo dialect of Mascioni, that we are about to examine, highlights the fact that in certain cases it is difficult to individuate two different functions for TV and inflection. In (26) the stress is indicated only in the case of the antepenultimate or final position. However, for the sake of clarity, the stressed vowel is indicated in bold. The grey color indicates the shapes of 1pl and 2pl, which in the imperfect combine two thematic elements.

(26)	Present indicative	Imperfect indicative	Imperfect subjunctive	Perfect
a.	j**ɔ**k-o	jok-**e**-a	jok-**e**-ss-e	joˈk-**ɔ**
	j**o**k-i	jok-**i**-i	jok-**i**-ʃti	jok-**i**-ʃti
	j**ɔ**k-a	jok-**e**-a	jok-**e**-ss-e	joˈk-**ɔ**
	jok-**e**-mo	jok-e-**a**-mo	jok-e-ss-**ɛ**-mmo	jok-**ɛ**-mmo
	jok-**e**-te	jok-e-**a**-te	jok-e-ss-**ɛ**-ʃte	jok-**ɛ**-ʃte
	ˈj**ɔ**k-enu	joˈk-**e**-enu	joˈk-**e**-ss-eru	jok-**ɔ**-ru
	'I play, etc.'	'I was playing, etc.'	'(if) I played, etc.'	'I played, etc.'
a'.	joˈk-**a** 'to play'			
b.	bb**e**j-o	bbej-**e**-a	bbej-**e**-ss-e	bb**e**bb-e
	bb**i**j-i	bbej-**i**-i	bbej-**i**-ʃti	bbej-**i**-ʃti
	bb**e**j-e	bbej-**e**-a	bbej-**e**-sse	bb**e**bb-e
	bbej-**e**-mo	bbej-e-**a**-mo	bbej-e-ss-**ɛ**-mmo	bbej-**ɛ**-mmo
	bbej-**e**-te	bbej-e-**a**-te	bbej-e-ss-**ɛ**-ʃte	bbej-**ɛ**-ʃte
	bb**i**j-u	bbeˈj-**e**-enu	bbeˈj-**e**-ss-eru	ˈbb**e**bb-eru
	'I drink, etc.'	'I was drinking, etc.'	'(if) I drank, etc.'	'I drank, etc.'
b'.	bbej-e 'to drink'			

c.	**ɔ**rm-o	orm-**e**-a	orm-**e**-ss-e	orˈm-**i**
	orm-i	orm-**i**-i	orm-**i**-ʃti	orm-**i**-ʃti
	ɔrm-e	orm-**e**-a	orm-**e**-ss-e	orˈm-**i**
	orm-**e**-mo	orm-e-**a**-mo	orm-e-ss-**ɛ**-mmo	orm-**ɛ**-mmo
	orm-**e**-te	orm-e-**a**-te	orm-e-ss-**ɛ**-ʃte	orm-**ɛ**-ʃte
	orm-u	orˈm-**e**-enu	orm-e-ss-eru	orm-**i**-ru
	'I sleep, etc.'	'I was sleeping, etc.'	'(if) I slept, etc.'	'I slept, etc.'

c'. orˈm-**i** 'to sleep'

Mascioni

The patterns discussed so far emerge in the data in (26) in an even more complex form: (i) The 1st and 2nd plural in the present indicative and the entire imperfect indicative show the specialized TV *-e-*; (ii) *-e-* is unstressed in the 1st and 2nd plural of the imperfect indicative (blackened), exactly as in the standard Italian forms like *ved-e-và-mo/ved-e-và-te* 'we/you saw' and in the 3rd plural of the present indicative of the first class, between the root and the inflection *-nu*, cf. *jɔk-enu* 'they play'; (iii) The ending *–enu* occurs also in the 3rd plural of the imperfect, suggesting that it can be treated as a single element; (iv) In the imperfect subjunctive *-e-* occurs in all persons except the 1st and the 2nd plural where, as in the imperfect indicative, it is unstressed and interpolated between the root and the inflectional string *–ss-ɛ-mmo/ -ss-ɛ-ʃte* (blackened), where the same stressed TV *-ɛ-* as in the perfect occurs; (v) Both weak and strong perfects present the same TV pattern in 2nd singular and 1st and 2nd plural, with the metaphonized outcome *-i-* from *-e-* before final *-i* in the 2nd singular and the open result *-ɛ-* in the two plural persons. The metaphonetic outcome *-i-* occurs also in the 2nd singular of the imperfect subjunctive and in the perfect; (vi) In weak perfects the 1st singular and 3rd singular and plural share the same stressed vowel. The first class introduces a specialized stressed morpheme *-ɔ* that shares with the TV the position and the stress, occurring also in the 3rd plural between the root and the ending *-ru*. The fourth class introduces *-i/ -i-ru*.

Schematizing, the TV *-e-* and its metaphonetic outcome *-i-* occur throughout the paradigm without class distinction. In the imperfect indicative *-e-* is the only realization of the tense/ aspectual property, while in the subjunctive it combines with the specialized inflection *-ss-*. Since in this dialect *-e-* appears also in the 1st and 2nd plural of the present indicative, we need to assume that it is not specialized for tense and aspect, as suggested in (27) where it is characterized as the exponent for the argumental variable.

(27) e_x

In the 1st and 2nd plural of the imperfect, *-e-* is followed by a stressed vowel preceding the inflectional elements, as in *jok-e-**a**-mo* 'we played'/*jok-e-ss-**ɛ**-mmo* '(if) we played'. The simplest hypothesis is to think that *-a-* and *-ɛ-* are in turn thematic vowels specialized for the imperfect, respectively indicative and subjunctive. They host the main stress as real TVs and precede the inflectional endings, as in (28). The amalgam created by (28a), with *-e-* as TV, is merged to the mood exponent *-ss-* in (28b), and this combination in turn to specialized TV *-ɛ-* in (28c). Finally (38d) yields the 1st plural inflected form.

(28)

a. $< jɔk_R, e_{TV} > \rightarrow [\ [jɔk_R]\ e_{TV}]$

b. $[_v < [[jɔk_R]\ e_{TV}],\ ss_{PastSubj} > \ldots \rightarrow [_v\ [\ [\ [jɔk_R]\ e_{TV}]\ ss_{PastSubj}] \ldots$

c. $[_v < [\ [jɔk_R]\ e_{TV}]\ ss_{PastSubj}],\ ɛ_{TV} > \rightarrow [_v\ [[[\ [jɔk_R]\ e_{TV}]\ ss_{PastSubj}]\ ɛ_{TV}] \ldots$

d. $[\ _T < [[[\ [jɔk_R]\ e_{TV}]\ ss_{PastSubj}]\ ɛ_{TV}],\ mmo_{Infl} > \rightarrow [_T\ [[[[[jɔk_R]\ e_{TV}]\ ss_{PastSubj}]\ ɛ_{TV}]mmo_{\varphi}] \ldots$

Thus, we have clear evidence of the fact that TVs are able to systematically interact with other types of inflectional exponents, and they can be duplicated as in (28c), by a TV specialized for the mood or tense inflection. The same applies to the imperfect indicative *-e-a-*, where *-a-* is the specialized morpheme for the 1st and 2nd plural persons, as in (29a), where the interpretation introduced by *-a-* can be specified as Past Progressive. The fact that *-ɛ-* also occurs in the perfect, cf. *jok-**ɛ**-ʃte* 'you played', suggests that it is connected to non-Discourse Linked contexts, in general sense Irrealis, as indicated in (29b).

(29)

a. a ←→ [x, Past Progressive]

b. ɛ ←→ [x, Irrealis]

The weak perfect introduces a specialized TV in the 1st singular and the 3rd singular and plural, *-ɔ-* and *-i-*. Again, it is difficult to classify *-ɔ-* as a TV or an inflectional exponent. However, its distribution, particularly its occurrence between the root and the inflection in the 3rd plural person, e.g., *jok-**ɔ**-ru* 'they played', identifies it as a tense/ aspectual specialized TV, as in (30).

(30) ɔ ←→ [x, Past]

3.4. Inflections and TVs as exponents of agreement features: inflection or TV?

Sharing the same inflectional structure in the 1st and 2nd person plural is a widespread pattern. A good example is given by the present indicative of the Piedmontese Franco-Provençal of Coazze (Turin), which presents a single inflectional and thematic paradigm throughout the verbal classes, as in (31). The levelling affects both person exponents (1st singular, 2nd singular and 3rd singular and plural) and thematic forms, *-ø-* and *-ɛi*. The exponent *–nt* includes 1st and 3rd plural. As already noticed for other systems, the SCls are partially syncretic, specifically 1st/3rd person plural/feminine, 2nd plural/3rd singular masculine.

(31)

a.	i tʃam-u	b.	i vɛj-u	c.	i kor-u	d.	i drøm-u
	t tʃam-e		t vɛj-e		t kor-e		t drøm-e
	u/i tʃam-at		u/i vɛj-at		u/i kor-at		u/i drøm-at
	tʃaˈm-ø-nt		viˈj-ø-nt		kuˈr-ø-nt		dryˈm-ø-nt
	u tʃaˈm-ɛi		u viˈj-ɛi		kuˈr-ɛi		u dryˈm-ɛi
	i tʃam-unt		i vɛj-unt		i kor-unt		i drøm-unt
	'I call, etc.'		'I see, etc'		'I run, etc.'		'I sleep, etc.'
a'	tʃaˈm-ɛ 'to call'	b'.	vere 'to see'	c'.	kore 'to run'	d'.	dryˈm-i 'to sleep'

This pattern exemplifies the other side of the phenomenon concerning the role of inflectional endings. In fact, in (31) personal inflections contribute to realizing the present as shown by comparison with other paradigms. In the latter, in fact, they are added to the suffix of tense/aspect, with the exception of the forms 1º and 2º plural, which introduce a specialized TV. The simplest idea is that forms like *tʃam-u* 'I call' are devoid of a tense/ aspect specification and their use is only based on the predicate nature of the root, with the result that the agreement element is sufficient to associate the predicate with the argument, in (31a), yielding a verb form able to realize T.

(31)
a. $< tʃam_R, u_\varphi > \rightarrow [_{v/\varphi} [tʃam_R] u]$
b. $[_T [_{v/\varphi} [tʃam_R] u] \ldots$

This means that TVs are selected by 1st and 2nd plural persons, as in (32a,b), or, more precisely, we can assign them the specialized content of 1st and 2nd plural person respectively:

(32)
a. $ø_{\text{1st plural}} \leftrightarrow / R \_\_$
b. $ɛi_{\text{2nd plural}} \leftrightarrow / R \_\_$

In other words, these elements can by now be associated with agreement features. Moreover, in the 1st plural the exponent *–nt* is joined, as in (33). We can treat that element as endowed with the only property [⊆] as far as it occurs in both the 3rd plural and in the 1st plural.

(33)
a. $< tʃam_R, ø_{TV} > \rightarrow [[tʃam_R] ø_{TV}]$
b. $[_T < [[tʃam_R] ø_{TV}], nt_{Infl} > \ldots \rightarrow [_T [[[tʃam_R] ø_{TV}] nt_{Infl}] \ldots$

This hypothesis is supported by the fact that thematic vowels display the same behavior as the other verbal inflectional elements, morphemes, and clitics. Specifically, they are sensitive to φ-features and person differences, thus showing:

- Syncretism
- Extension to sub-sets of persons
- Complementary distribution with inflectional elements

4. Some final remarks

The data we have looked at shows some specific patterns that emerge in the paradigms of different varieties. A typical pattern concerns the 1st and 2nd plural persons, which generally require a special expression, distancing themselves from singular and 3rd plural morphology. This pattern is highlighted in (26) and in (31), while characterizing only partially the Northern dialects examined in the preceding sections (see the first plural in (5)). It is present in Romance varieties in different ways (cf. Maiden 2018), and in Standard Italian appears in the first plural present indicative and in the imperfect paradigms.

The best-known property of 1st and 2nd plural persons is that they include the deictic reference to one of the interlocutors and the reference to other individuals. The latter involves the third person reference, that is, anaphoric with respect to discourse. We see that this composed interpretation is associated in many cases with a specialized way to introduce the reference to the subject, an evident kind of Differential Subject Marking, including either special generalized inflections or the duplication of the TV, as in *jok-**e**-**a**-mo* 'we played'/*jok-**e**-ss-**ɛ**-mmo* '(if) we played' discussed in (28) and (29), as well as in the Standard Italian forms *gioc-**a**-v-**a**-mo* 'we played'. The latter solution, i.e., doubling of TV, can be treated as a way of expressing the composed reference. In other words, the TV associated with the root is integrated with a TV connected with the tense/aspect or mood specification. The generalization of the same morphology for the 1st and 2nd plural persons, as in (32), is based on a specialized TV. Again, the TV expresses the same reference as the subject. Naturally, the syntactic computation combines sub-word elements based on their selection restrictions. We could think that the observed grammatical tendencies reflect the role of semantic and conceptual properties, at least partially governed by restrictions not specific to the language. In other words, selection restrictions are the result of the way references and events are conceptualized. A similar suggestion can be extended to the issue, that here we only mention, of the mutual order of inflectional suffixes.

Summing up, our discussion has faced some theoretical and descriptive points concerning morpho-syntactic phenomena involving the internal structure of words. Syncretism and the role of TV in Romance languages provided the testing ground for the morphological level of analysis. In contrast to the split between syntactic structures and morphological representations pursued by DM, which confines morphology to an extern level, a more significant and non-ad-hoc treatment is available, based on the idea that no autonomous morphological component is necessary or useful. We have assumed that all morphological elements are endowed with interpretable content, including inflectional exponents and that Merge and amalgamation operations are able to account for the formation of complex words on principled bases.

References

Anderson, S. 1992, *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, M. 1994. *Morphology by Itself*. Cambridge, Mass.: MIT Press.

Aronoff, M. 1998. Isomorphism and monotonicity. In S. Lapointe, and D. Brentari (eds.), *Morphology and Its Relation to Phonology and Syntax*, 411–418. Stanford: CA: CSLI,.

Baker, M. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Baldi, B., and L. M. Savoia. 2021. Partitives and Indefinites: Phenomena in Italian Varieties. *Studia Linguistica:* 1–45. Online.

Belvin, R., and M. Den Dikken. 1997. There, happens, to, be, have. *Lingua* 101: 151–183.

Calabrese, A. 2015. Irregular Morphology and Athematic Verbs in Italo-Romance. *Isogloss*: 69–102.

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

Chomsky, N. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36(1): 1–22.

Chomsky, N. 2015. Problems of projections. In E. Di Domenico, C. Hamann, and S. Matteini (eds.), *Structures, Strategies and Beyond. Studies in honour of Adriana Belletti*, 3–16. John Benjamins Publishing Company: Amsterdam/Philadelphia.

Chomsky, N. 2020a. *The UCLA Lectures* (April 29 – May 2, 2019). https://ling.auf.net/lingbuzz/005485

Chomsky, N. 2020b. Puzzles about Phases. In L. Franco and P. Lorusso (eds.), *Linguistic Variation: Structure and Interpretation: Contributions in Honor of M. Rita Manzini*, 163–167. Berlin: De Gruyter.

Chomsky, N. 2021. Minimalism: Where Are We Now, and Where Can We Hope to Go, *Gengo Kenkyu* 160: 1–41.

Chomsky, N., Gallego, Á. 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*, *Catalan Journal of Linguistics*. *Special issue*, 229–261.

Collins, Ch., and R. Kayne. 2020. *Towards a Theory of Morphology as Syntax*, online, http://ordinaryworkinggrammarian.blogspot.com/2020/12/towards-theory-of-morphology-as-syntax.html (13.07.2021).

Embick, D. 2010. *Localism versus Globalism in Morphology and Phonology*. Cambridge. Mass.: The MIT Press.

Embick, D., and R. Noyer. 2001. Movement Operations after Syntax. *Linguistic Inquiry* 32(4): 555–595

Franco, L., Manzini, M. R., and L.M. Savoia. 2020. Locative Ps as general relators: Location, direction, DOM in Romance. *Linguistic Variation*, online https://doi.org/10.1075/lv.00035.fra, 2–38.

Halle, M., and A. Marantz. 1993. Distributed morphology and the pieces of inflection. In K. Hale, and S. J. Keyser (eds.), *The view from Building 20*, 111–176. Cambridge Mass-: The MIT Press.

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*, 275–288. *MIT Working Papers in Linguistics 21*.

Hauser, M. D., Chomsky, N., and W. T. Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298: 1569–1579.

Manzini, M. R. 2021. Chomsky's (2020) Links and linker phenomena. *Quaderni di Linguistica e Studi Orientali QULSO* online 7: 89–102.

Manzini, M. R., and L. M. Savoia. 2005. *I dialetti italiani. Morfosintassi generativa*. 3 vv., Alessandria: 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. 2011a. *Grammatical Categories: Variation in Romance Languages*. Cambridge: Cambridge University Press.

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

Manzini, M. R., and L. M. Savoia. 2017. Gender, number and inflectional class in Romance: Feminine/plural *–a*. In J. Emonds, and M. Janebova (eds.), *Language Use and Linguistic Structure*, 263–282. Olomouc: Palacky University Olomouc.

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

Manzini, M. R., Savoia, L. M., and B. Baldi. 2020. Microvariation and macrocategories: Differential Plural Marking and Phase theory. *L'Italia Dialettale* 81: 189–212.

Marantz, A. 2001. *Words*. WCCFL XX Handout, USC, February 2001

Marantz, A. 2007. Phases and Words. In S.-H. Choe (ed.), *Phases in the theory of grammar,* 191–222. Seoul: Dong-In Publishing Co.

Maiden, M. 2012. *The Romance Verb. Morphomic Structure and Diachrony*. Oxford: Oxford University Press.

Nyrop, K. 1903. *Grammaire historique de la langue française.* Copenhague: Det Nordiske Forlag.

Renzi, L., and L. Vanelli. 1983. I pronomi soggetto in alcune varietà romanze. In AA.VV., *Scritti linguistici in onore di G.B. Pellegrini*, 120–145. Pisa: Pacini.

Roberts, I. 2010. *Agreement and Head Movement*. Cambridge, Mass.: The MIT Press.

Roberts, I. 2018. Object clitics for subject clitics in Francoprovençal and Piedmontese. In M. Grimaldi, R. Lai, L. Franco, and B. Baldi (eds.), *Structuring Variation in Romance Linguistics and Beyond*, 255–265. Amsterdam: Benjamins.

Savoia, L. M., Baldi, B., 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.

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

Thornton, A. 2005. *Morfologia*. Roma: Carocci.

Villanueva Svensson, M. 2021. The origin of the Indo-European simple thematic presents. The nominal connection. *Indo-European Linguistics*: 1–29.

Wood, J., and A. Marantz. 2017. The interpretation of external arguments. In R. D'Alessandro, I. Franco, and Á. J. Gallego (eds.), *The Verbal Domain*, 255–278. Oxford: Oxford University Press.

LingBaW *Linguistics Beyond And Within 8 (2022), 24-36*

Accent boundaries and linguistic continua in the laryngeal subsystems of English

Katalin Balogné Bérces
Pázmány Péter Catholic University, Budapest, Hungary
Catholic University in Ružomberok, Slovakia

Abstract
A parallel is drawn between the northernmost regions of England represented by Durham and Yorkshire and the transition zone Ouddeken (2016) identifies between voicing and aspiration languages in the Dutch-German dialect continuum. It is argued that, owing to historical changes and dialect contact, the Northern Englishes discussed exhibit hybrid laryngeal systems as a result of being geographically intermediate between Scots in Scotland, which is a voice language similar to Dutch, and mainstream varieties of English spoken more to the south in England (and in most of the rest of the English-speaking world), which are aspiration systems of the German type. We model the emergence of laryngeal systems as the setting of three parameters: (i) whether the laryngeally marked/specified obstruent series contains [voice] (L-system) or [asp] (H-system); (ii) whether the laryngeal prime is able to spread (right-to-left); and (iii) whether the system has pre-obstruent delaryngealisation (POD) (due to which in C1C2, C1 becomes unmarked/underspecified). While spreading L with POD derives voice languages and non-spreading H with no POD derives aspiration languages, two mixed combinations derive the intermediate categories of Durham and Yorkshire (spreading L & no POD and spreading H & no POD, respectively). We also show that all remaining combinations are attested cross-linguistically or else theoretically uninterpretable.

Keywords: laryngeal phonology, laryngeal typology, accents of English, laryngeal realism, voice assimilation

1. Background

Languages vary as to how many series of obstruents they distinguish by some laryngeal specification, such as voicing, aspiration (or spread glottis), glottalisation (constricted glottis). Those with a single series realise obstruents as tenuis, i.e., voiceless unaspirated unglottalised (e.g., Hawaiian, Maori), and it has been observed that this series is attested in all other, more complex systems, i.e., languages with two, three or more sets of obstruents. In the following discussion, we narrow the focus down along two dimensions of the existing typology. *First*, we concentrate on binary laryngeal obstruent systems only: languages and varieties in which obstruents are organised into two contrasting sets in terms of laryngeal activity. *Second*, we limit our scope to contrasts expressible in terms of the acoustic measure of Voice Onset Time

(VOT), in line with Lisker and Abramson's (1964) seminal proposal that this measure can serve to capture, in a unitary mode, voicing categories of stops across languages which had previously been distinguished by the seemingly unrelated phonetic features of voicing, aspiration and fortisness/lenisness – as a result, glottalised forms will henceforth be ignored. The remaining categories, i.e., VOT-based binary systems, subsume most languages spoken in Europe, and will therefore perfectly suit our purposes.

In this framework, which is commonly referred to as laryngeal realism (because it is more realistic than the traditional approach solely based on [±voice] and not considering VOT differences – see Honeybone 2005; Iverson and Salmons 2008; etc.), two language types are differentiated. On the one hand, we attest (true) voice (or voicing) systems, which base the laryngeal contrast between the two series of obstruents on a voiced~voiceless distinction (e.g., [b]~[p]), i.e., underlying phonological [±voice] is phonetically realised by (pre-)voiced vs. tenuis. Slavic and Romance languages are typically taken to belong to this category. Other languages, on the other hand, place the functional load of expressing the contrast on aspiration in the form of the distinction between aspirated~unaspirated (or fortis~lenis [in terms of articulatory force], e.g., [pʰ]~[b̥]), i.e., [±voice] is manifested by spread glottis (positive VOT) vs. (variably partially voiced) tenuis. These are called aspiration systems, and mainstream varieties of English (typically simply labelled "English") and German are unambiguously classified as such.

According to a widely-held view, voice and aspiration languages do not only differ in the phonetic forms of their plosives, but also in the phonological patterning of their whole obstruent system. In particular, it is only voice systems in which obstruents exhibit regressive voice assimilation (RVA): a voiced obstruent such as [b] is (fully) devoiced when it is followed by a voiceless obstruent such as [t] (and the other way round), deriving [-pt-] in words like *obtenir* in French. In contrast, in most accents of English *obtain* retains [-b̥tʰ-], and in general, claims have been made that in aspiration systems no systematic, phonological laryngeal spreading is attested (see Balogné Bérces and Huszthy 2018; Huber and Balogné Bérces 2010).

It is of crucial relevance to the present discussion that historical changes can induce a switch from one category to another – as a result, languages belonging to the same language family do not necessarily belong to the same laryngeal type. Within Germanic, North Germanic languages as well as most varieties of English and German, North-Eastern dialects of Dutch are aspiration systems – faithfully reflecting their Germanic ancestry. However, a number of their sisters (and their descendants), under the historical contact influence of Romance or Slavic, replaced the original aspiration system with one resting on voice, so much so that they even integrated RVA into their phonological systems. As a consequence, present-day Yiddish, (Western and Southern) Dutch, Afrikaans, (West) Frisian, Rhineland German are voice languages with RVA (see Jansen 2004 and references therein). Dutch is frequently described as possessing a split obstruent system, in which fricatives behave as those of aspiration languages but plosives spread [voice] (see, e.g., Iverson and Salmons 2003) – as this issue is irrelevant to our argumentation, it will not be considered, and (standard) Dutch will henceforth be simply regarded as a voice system.

A final point to be made concerns the label "English": it should be clear by now that, due to its extensive accent variation, it cannot be treated as a uniform laryngeal class but its

mainstream varieties (which we dub General English (GE); i.e., the most well-known ones, incl. the standard accents) need to be treated separately from certain other dialects. Specifically, while GE can be considered the epitome of aspiration systems, already Iverson and Salmons (1999) notes that Scots/Scottish English belongs to the category of voice languages: its fortis plosives are unaspirated, its lenis set has true voicing, and it is reportedly characterised by RVA (see below). The present paper claims that the distinction between aspiration varieties and voice varieties of English is crucial for the understanding of the laryngeal settings attested in certain North-of-England varieties (also see Balogné Bérces to appear).

We have seen, then, that in laryngeal phonological typology, standard forms of Dutch are classified as voice systems, due to phonetic voicing in their lenis obstruents and the lack of aspiration in their fortis counterparts – unlike most other Germanic languages such as Standard German, which is characterised by aspirated voiceless plosives for fortis and (variably) voiceless unaspirated (tenuis) for the lenis set. However, as Ouddeken (2016) points out, the transition zone of the Dutch-German dialect continuum in Europe comprises intermediate systems with a phonetic overlap between VOT values for fortis and lenis plosives. In this paper we claim that the situation Ouddeken sketches out bears a close resemblance to the present-day distribution of laryngeal systems in Britain brought about by historical changes and dialect contact: a number of regions in the far north of England, sandwiched between Scots (a voice language, as it will be recalled, with RVA) and mainstream varieties of English English spoken to the south (belonging to the type referred to above as GE, an aspiration language), have been reported to exhibit hybrid laryngeal systems that may lack aspiration and have partial, asymmetrical voice assimilation. We will show below how these northern English accents can be modelled as mixed (or fudged) lects, using the tools of theoretical laryngeal phonology.

The paper is structured as follows: Section 2 introduces the findings of Ouddeken's (2016) acoustic corpus study of the phonetic and phonological properties of laryngeal systems in the transition zone between (voicing) Dutch and (aspirating) German, then Section 3 describes similar hybrid varieties spoken in the north of England to argue that they have emerged as a reaction to an analogous situation of transition. Section 4 sketches out the theoretical devices applied and how they model "plain" forms of voice and aspirating languages, as well as the ways parameter settings account for the combinations of elements of laryngeal phonological patterning. Finally, Section 5 concludes.

2. The transition zone in the Dutch-German dialect continuum

Ouddeken (2016) reports the results of a study carried out on VOT values in the varieties spoken in the Dutch-German dialect continuum, which proves to be an ideal testing ground for accent contact since in these regions both types of laryngeal system are present in such a way that the standard language is Dutch (a voice language) for some speakers, and German (an aspiration system) for others. Ouddeken presents VOT measurements of word-initial plosives on the one hand, and percentages of voicing during closure in plosive clusters (i.e., in the assimilation context) on the other, with data retrieved from different databases.

The results reveal a continuum of both variables investigated, in which the western end (geographical longitude of cca. 5-7°) constitutes an unambiguous case for a voice system and the eastern end (geographical longitude of cca. 9-11°) exemplifies the aspiration system. Between these two ends, from west to east, however, a gradual transition is observed for both variables. As for VOT values, there is a gradual increase for each plosive: lower values are found in the west, higher values in the east (ranging, roughly, from -100 msec to 20 msec for lenis, and from cca. 20 msec to over 40 msec for fortis; of course, there is a visible place-of-articulation effect causing some variability). What Ouddeken calls the middle area exhibits hybrid systems with both prevoicing and positive VOT's, and with a huge amount of variation in the data. She concludes that this transition zone is characterised by phonetic overlap, but one in which most individual speakers still make a distinction between the two series.

As for the assimilation data, the same kind of continuum can be identified. While the westernmost areas exhibit stable cases for voice assimilation (100% voicing during closure in lenis-final clusters and close to zero in fortis-final ones) and the easternmost regions display virtually no assimilation (with quite some dispersion in the data, due to the differences in the input clusters), the same middle area can be identified as above, showing a hybrid pattern of RVA. Here, both fortis-final and lenis-final cluster types can show full intervocalic voicing (i.e., there are fully voiced clusters attested even with a fortis C2), but this is inconsistent in both cases, and again, variation is extensive. What this indicates is that in the transition zone both the aspiration and the voice feature seem to be, albeit in an inconsistent manner, phonologically active, and she concludes that for systems where plosive clusters undergo full intervocalic voicing, it has to be assumed that neither feature is present – what appears to be voicing is in fact a surface phonetic process (intervocalic passive voicing of obstruents unmarked lexically for laryngeal properties).

Whereas most of Ouddeken's argumentation and its theoretical implications are irrelevant to our discussion (due to the fact that she primarily (only?) focuses in her conclusions on which laryngeal prime is active/present phonologically while, as we will see below, the present paper takes a different direction, leading to a more sophisticated model[1]), it is worthy of note that the transition zone between pure aspiration and voice systems exhibits graduality and fuzziness as well as a great deal of variation in both the phonetic realisation of plosives and the patterning of RVA. In what follows we argue that a similar situation of transition has led to the emergence of hybrid laryngeal systems in a middle area between the voice system of Scots in Scotland and the aspiration systems of English in England, with variable phonetic realisations of obstruents, and voicedness and voicelessness being variably, asymmetrically active.

[1] Actually, this difference may have some bearing on the fact that Ouddeken's middle areas seem to show the characteristics of mixed lects (Chambers and Trudgill 1998/2004: 110–113) as she extensively argues in Ouddeken (2018), with the transition varieties mixing the variable's realisations found in the non-transition varieties; however, the Northern English cases discussed below will eventually turn out to be fudged lects, i.e., by mixing (phonological) components of the variable they produce (systematically used) novel realisations (assimilation systems unattested outside the transition zone). How Chambers and Trudgill's classification applies to the phenomena at hand needs further investigation.

3. Hybrid laryngeal systems in varieties of Northern English

As mentioned previously, Germanic languages faithfully reflecting their historical ancestry belong to the aspirating laryngeal type. This makes Scots a surprising odd language out: there are reasons to assume that already in Older Scots, voiceless stops were unaspirated and lenis plosives were (fully) voiced (Johnston 1997). How it had developed into this system is unclear (crucially, for instance, Celtic languages are also typically aspiration systems so a simple Celtic substrate effect is difficult to posit); however, this is a firmly established, well-documented property of the language. All the more recent (and phonetically more reliably accurate) descriptions report that in Modern and Present-day Scots (as well as Scottish Standard English) fortis plosives tend to be unaspirated (or at least more weakly aspirated than in GE) while the lenis series is (pre-)voiced (except perhaps for speakers from the Central Belt – with the urban centres of Glasgow and Edinburgh – only, and perhaps with shorter fortis VOT in the east than in the west and for older and working-class speakers than for younger and middle-class speakers).[2] In this respect, Scottish English/Scots seems to represent the same category as (general) Romance or Slavic – or (standard) Dutch (see Iverson and Salmons 1999: 22–23).

In addition, Scots also resembles Dutch more than GE in exhibiting RVA: Abercrombie (1967: 135–136) discusses this process explicitly, stating that it is "found very commonly, though not universally, among speakers of educated Scots" (Abercrombie 1967: 136), and giving the examples *blackboard* with [-gb-], and *with them* and *birthday* with [-ð-]. To this Wells (1982: 412–413) adds the example *mos(t) valuable* with [-zv-].[3]

Most of the rest of Britain (and the British Isles in general), however, houses varieties belonging to GE: they are aspiration systems, in which the aspirated/fortis set is stably voiceless, the unaspirated/lenis set is realised as tenuis in most positions and undergoes word-internal and cross-word passive voicing, i.e., they assume the voicing of surrounding sonorants to a highly variable degree. Therefore, whereas in utterance-edge (GE *cheese* [-z̥]) and pre/post-fortis (GE *cheesecake* [-z̥-]) positions they tend to be partly voiced or voiceless, in sonorant contexts they vary from partly to fully voiced (GE *cheeses* [-z-]). The few examples of accents that diverge from this pattern are found in the (far) north of England. For some of these North-of-England varieties descriptions are scarce in detail so they need to be corroborated: e.g., Black Country English voiced initial and final consonants are reported to be fully voiced, and there appears to be some written evidence for final devoicing in Birmingham (Clark 2004). The two most well-documented cases are provided by the areal phenomenon commonly referred to as "Yorkshire Assimilation" and the dialect of English in Durham. As we will see, these cases cannot be considered forms of GE since they feature RVA. Yet, they are not straightforward representatives of voice systems, either: their RVA is partial or asymmetrical (voicelessness-spreading or voicedness-spreading only), and some of them may have more aspiration than Scots or Dutch. In what follows, we introduce these two cases,

---

[2] See Wells (1982: 409), Masuya (1997), Stuart-Smith (2004), Scobbie (2005, 2006), Watt and Yurkova (2007), Docherty et al. (2011), Stuart-Smith et al. (2015), Sonderegger et al. (2020), etc.

[3] For a detailed discussion of a problematic aspect of the data, see Balogné Bérces to appear.

and argue that they are hybrid systems very much like the ones Ouddeken identifies in the Dutch-German transition zone and for very much the same reasons.

Yorkshire Assimilation, as the dialectological literature (Wells 1982: 366–367; Trudgill 1999: 70) reports, characterises certain North-of-England English varieties; it has been attested in several parts of Yorkshire (especially West and South Yorkshire), which has granted it the label it is referred to with. It is a kind of devoicing assimilation, i.e., it only affects voiced/lenis obstruents that stand before a voiceless/fortis one (see (1) below), but descriptions are unclear on which obstruents it involves. In certain areas, West Yorkshire in particular, it seems to be restricted to plosives – see esp. Hughes et al. (2012: 106), Wilmhelm (2018), Whisker-Taylor and Clark (2019).[4] In addition, almost all descriptions mention t-glottalling further affecting the [t] that derives from underlying /d/, leading to outputs like [ˈbraʔfəd] for *Bradford*, which thus appears to be a general auxiliary process.

From the perspective of Yorkshire Assimilation as a phonological process, its most relevant property is the phonetic forms of the resulting devoiced obstruents. Whisker-Taylor and Clark (2019) confirm Firth's (1991) acoustic findings that lenis obstruents like /b/ are realised in RVA contexts as [p] rather than [b̥], i.e., they do not carry any voicing whatsoever, unlike obstruents in the same phonological environment in GE. Therefore, it is straightforward that a merger takes place of underlying fortis and devoiced lenis (also supported by the fact that phonetic [t] undergoes glottalling irrespective of its source): this is a case of voicelessness spreading into the preceding obstruent. See some illustration in (1).

(1) Yorkshire assimilation[5]

jazz	[-z̥]	*pass*	[-s]
jazz music	[-z̥m-]	*pass Molly*	[-sm-]
jazz band	[-z̥b̥-]	*pass Barry*	[-sb̥-]
jazz dance	[-z̥d̥-]	*pass Dave*	[-sd̥-]
jazz club	[-skʰ-]	*pass Keith*	[-skʰ-]
jazz pub	[-spʰ-]	*pass Pete*	[-spʰ-]

As is shown in the examples, the pre-pausal and pre-sonorant positions as well as the obstruent+lenis sequences (*jazz band* and *pass Barry*) surface in the same form as in GE (although the final two would merge in [-z̥b̥-] or [-zb-] in true voice systems or – as argued below – in Durham), whereas the pre-fortis case presents a neutralising environment in which *jazz club* collapses with *pass Keith* in [-skʰ-], unlike in GE, where lenis+fortis sequences surface unmodified, as [-z̥kʰ-].

Durham English (more precisely, "the low-status Durham Vernacular" – Kerswill (1987: 42), on the other hand, contains fully voiced lenis and voiceless unaspirated (tenuis) fortis obstruents, which engage in a kind of voicing (i.e., voicedness) assimilation that purportedly spreads [+voice] only. In all the examples provided in the descriptions (of which a sample is

[4] Whisker-Taylor and Clark (2019) constitutes the first systematic empirical analysis of Yorkshire Assimilation, reporting data from Huddersfield, West Yorkshire.

[5] The data have been adapted from Honeybone (2011). For certain varieties exhibiting Yorkshire Assimilation, the diacritic for devoicing and the superscript [ʰ] for aspiration may not be (fully) justified. These phonetic facts need empirical verification but do not disturb the argumentation here.

given in (2) – see Kerswill 1987: 42, 44; Harris 1994: 137–138; Cyran 2014), a fortis+lenis sequence surfaces as lenis+lenis, and apparently the process affects plosives, fricatives and affricates alike.

(2) Durham assimilation

top gun	[-bg-]	*pitch black*	[-dʒb-]	*scraped down*	[-bdd-]
football	[-db-]	*each deputy*	[-dʒd-]	*what's gone*	[-dzg-]
backbone	[-gb-]	*this village*	[-zv-]		

Note how the cluster in *football* is claimed to become identical with that of hypothetical "foodball" – the fully voiced final obstruent in these sequences causes all preceding obstruents to assume full voicing. By analogy with *this village* with [-zv-], then, we expect [-zb-] in *pass Barry* from (1) above. This part of the phenomenon makes Durham English resemble true RVA-systems; however, its lenis obstruents cannot lose their voicedness and become either tenuis or fortis: *jazz club*, which would surface with [-sk-] in a voice language, retains voicing in the [z] in Durham, very much as in GE.

An additional property of Durham English is constituted by cross-word pre-sonorant voicing affecting the voiceless obstruents: e.g., Kerswill (1987: 44) reports examples like *like me* with [-gm-] as the cross-word cluster in the pronunciation. This phenomenon of sandhi-voicing is relatively well-attested, primarily in languages with clearly active voicing and final obstruent devoicing, from Sanskrit through Slovak, Western Dunántúl Hungarian and West Flemish to certain dialects of Breton, Catalan, Spanish, Italian and German (see Cyran 2012 and references therein). Perhaps the case that is the most extensively discussed in the phonological literature is that of Cracow Polish, in connection with which the process is indeed frequently referred to as "Cracow Voicing". Similarly to Durham English, a word like *brak* 'lack' changes its /k/ to [g] in pre-sonorant position, e.g., *brak oceny* 'lack of mark', *brak jasności* 'lack of clarity' (Cyran 2012: 154). Such cases are explained by Cyran (2012, 2014, 2017) as the passive voicing of an unmarked (i.e., tenuis) obstruent, which also seems to apply in Durham (although Cyran (2014), basing his analysis on somewhat different, more abstract theoretical considerations, proposes to treat Durham English as an aspiration system). Whatever the analysis, the fact that Durham English exhibits this voicing process that only applies across word boundaries (while there are no reports of it having final devoicing) lends further support to the claim that it is the manifestation of an unexpected, hybrid phonological system.[6]

4. Modelling the laryngeal subsystems of English

Recall from Section 1 that the theoretical framework the present discussion is couched in is laryngeal realism (and is, admittedly, also influenced by the assumptions of related Laryngeal

[6] One of the reviewers advises to check empirically how robust this sandhi voicing really is. Indeed, the examples cited (and re-cited) in the literature may be non-systematic (and are scarce in number anyway). I am aware that the whole system of Durham assimilation needs corroboration from empirical data systematically collected and phonetically analysed, and in fact this is part of the planned next phase of this research project.

Relativism, named so by Cyran (e.g., 2014)). A theoretical tool of crucial significance is laryngeal underspecification: in binary systems, the unmarked set is "unmarked" even in this sense (i.e., tenuis obstruents lack a laryngeal prime in the melodic representation, and they receive default interpretation in the phonetics), while the marked set contains [voice] (henceforth symbolised by L of Government Phonology / Element Theory – Harris 1994; Harris and Lindsey 1995; Backley 2011) in voice languages, and [spread glottis]/[asp] (henceforth: H) in aspiration languages. Therefore, in a voice system like Scots the (pre-)voiced vs. tenuis contrast (e.g., [b]~[p]) is the manifestation of phonological L vs. zero, whereas an aspiration system like General English with aspirated vs. (variably partially voiced) tenuis (e.g., [pʰ]~[b̥]), is based on H vs. zero.[7] In this latter, H-system, the optional and variable voicedness of tenuis is a result of what is called (phonetic) passive voicing that can affect unmarked/unspecified obstruents (and the prediction is that this only happens in aspiration languages – see Cyran 2014). Further support for the two opposite asymmetries between marked and unmarked comes from language acquisition research, which has shown that in voicing languages children acquire fortis plosives earlier than lenis ones, while in aspiration languages the chronology is the reverse (see Kager et al. 2007 and references therein).

It was also mentioned in Section 1 above that the typological difference between voice and aspiration systems is to a great extent (if not primarily) phonological: typically, in voice languages the [voice] feature (L) is phonologically active, causing symmetrical (both voicing and devoicing) RVA; in aspiration languages often no signs of any laryngeal activity are detectable. In GE, for instance, the fortis set is stably voiceless (and aspirated), and the lenis series is voiceless unaspirated (tenuis) and undergoes (word-internal and cross-word) passive voicing. Upon the concatenation of morphemes, whenever C1C2 obstruent sequences arise, both consonants preserve the phonetics attested in the citation form, no laryngeal spreading takes place: e.g., GE *ma<u>tch</u>* [-tʃ] + *<u>b</u>ox* [b̥-] yields *ma<u>tchb</u>ox* [-tʃb̥-] (cf. Hungarian (a voice system) *ma<u>tchb</u>ox* [-dʒb-] 'small toy car'); or GE *o<u>bt</u>ain* [-b̥tʰ-] (cf. French *o<u>bt</u>enir* [-pt-]). It is not clear whether this is due to the inability of the prime (H in this case) to spread, or the total absence of a laryngeal prime (previously proposed in Balogné Bérces and Huszthy 2018 and Huber and Balogné Bérces 2010), therefore for the present purposes we assume that the (in)ability of the laryngeal element to spread is a function of a language-specific parameter setting.

In a system like that of GE the only case when lenis/unmarked obstruents are considerably voiced phonetically is the intersonorant context, cf. GE *chee<u>s</u>e* [-z̥] and *chee<u>s</u>ecake* [-z̥-] vs. *chee<u>s</u>es* [-z-].

In contrast, in voice systems like Hungarian or (Warsaw) Polish, symmetrical RVA derives from the ability of the prime to spread (discussed above) on the one hand, and pre-obstruent delaryngealisation (POD) on the other. POD leads to the neutralisation of laryngeal contrast of C1 (the pre-obstruent consonant) in C1C2: it produces unmarked obstruents in C1, which are ready to receive laryngeal spreading from C2 (as in Hungarian *ma<u>tch</u>box* [-dʒ-]

[7] Note that this interpretation of [asp]/[voice] (i.e., H/L) has been present in Government Phonology/Element Theory since (at least) Harris (1994).

'small toy car' above); if, however, C2 is also unmarked, no spreading can happen, therefore both remain unmarked and will be phonetically interpreted by default (as in Hungarian *roadshow* [-t-] 'ibid.').

The chart in (3) below illustrates how the three phonological mechanisms introduced above function as three independent and freely combinable parameters: (i) whether the laryngeally marked/specified obstruent series contains L or H; (ii) whether the laryngeal prime is able to spread (right-to-left); and (iii) whether the system has POD. The two ends of a scale-like classification comprise voice languages (like Scots or Hungarian) with spreading L and POD (3a) and aspiration languages (like GE) emerging from non-spreading H accompanied by no POD (3d). These two systems diverge with respect to all the three parameters; however, the chart also shows that intermediate categories are also possible, and, although it may seem from the previous paragraph that POD always implies the ability to spread so that the two coexist within a system, this is hardly the case.

(3) Laryngeal systems derived from parameter settings

	a. symmetrical RVA	**b. voicedness-only RVA**	**c. voicelessness-only RVA**	**d. inactive laryngeal prime**
(i)	**L**	**L**	**H**	**H**
(ii)	**spreading**	**spreading**	**spreading**	**no spreading**
(iii)	**POD**	**no POD**	**no POD**	**no POD**
	voice lang.	**Durham**	**Yorkshire**	**asp. lang.**

As (3b) and (3c) respectively illustrate, two mixed combinations derive the intermediate categories of Durham (spreading L with no POD) and Yorkshire (spreading H with no POD). That is because the absence of POD ensures the stability of marked obstruents in the C1 position: L-marked lenis in Durham (preventing what would appear to be "voicelessness-spreading"), and H-marked fortis in Yorkshire (preventing apparent "voicedness-spreading"). When, however, C1 is occupied by a consonant of the unmarked series but C2 is marked, L-spreading (voicing RVA) happens in Durham and H-spreading (devoicing/fortisising RVA) happens in Yorkshire.

All remaining combinations of the settings of the three parameters are attested cross-linguistically or else theoretically uninterpretable. The H-system equivalent of voice languages in (3a) (i.e., spreading H with POD) is what Cyran (2012, 2014) takes to be the characterisation of the Cracow dialect of Polish. When spreading does not accompany POD, we get systems neutralising laryngeal features in pre-obstruent position: in such an L-system, all C1's in C1C2 end up as (unmarked) fortis irrespective of the laryngeal setting of C2[8]; in such an H-system, all C1's in an obstruent sequence end up as (unmarked) lenis (or at least, as de-H-ed, i.e., deaspirated).[9]

[8] It may well be the case that German is such a system: both *Jagden* [ˈja:kdən] 'hunt-PL' and *jagten* [ˈja:ktən] 'we/they hunted' surface with fortis C1 (see more data in, e.g., Piroth 2003) (and also note that German has word-final devoicing, which is a simple form of final lenition if the language is assumed to be an L-system).

[9] Such deaspiration is found in, e.g., Assamese (see, e.g., Dutta and Kenstowicz 2018), although more investigation is needed for a proper analysis as the language has a more complex, four-way laryngeal system.

When neither spreading nor POD is part of the phonology, it becomes untestable what the laryngeal prime present in the system is, and whether it is present at all – unless laryngeal realism is satisfied with relying on the sheer phonetic realisations of plosives, blindly assigning L to fully voiced/prevoiced and H to aspirated. A phonologically-based approach like the one pursued in this paper, however, will need evidence from phonological patterning (such as RVA), especially if we accept Cyran's principle in Laryngeal Relativism that the phonological prime receives phonetic interpretation arbitrarily (e.g., the H of Cracow Polish is realised as voiceless unaspirated, while its unmarked obstruents surface as voiced). With such assumptions made, the framework is unable to interpret the difference between the non-L-spreading system and the non-H-spreading one when POD is not effective – either can be considered to be the representation of, e.g., General English, which raises the theoretical issues of whether there is a laryngeal prime present in such a system at all (mentioned above) and whether it is necessary at all to assume two laryngeal elements, H and L, separately. These issues are, however, beyond the scope of the present discussion.

5. Conclusion

The discussion above has shown that varieties of Northern English like the ones spoken in parts of Yorkshire and Durham represent laryngeal systems intermediate between voice and aspiration languages. Their deviation from the mainstream, GE pattern has been previously noted, partly in (sometimes rather sketchy and anecdotal) dialectological descriptions, partly in the literature on laryngeal typology, and in passing elsewhere.[10] Previous work in laryngeal realism has also asserted that Scots is to be classified as a voice language, but made no closer examination of the dialectal variation in English English and the potential connection between the two phenomena.

In contrast, it has been stressed above that the geographical distribution of laryngeally deviant accents of English in the linguistic north is not considered accidental at all: they are all located in the transition zone between Scots-speaking regions and the GE-speaking rest of England. In a sense, their emergence should not be surprising since, as Ouddeken (2016) demonstrates, hybrid systems naturally arise under laryngeal contact – even though for historical and political reasons, the Scots-English dialect/language continuum does not exhibit the usual form of gradual transition in every aspect.

The linguistic links between Scots and (far) northern English are in fact well-known and widely discussed (see esp. Maguire 2012, Section 6; Honeybone and Maguire 2020, Section 3). Several pronunciation features fully or partially shared by Scots/Scottish English and northern English have been identified (e.g., Aitken's Law; certain pre-Great Vowel Shift vowels; the FOOT-GOOSE merger; STRUT [ʌ]; the retention of a /ʍ/ phoneme; etc.), and parallels are also found in grammar and lexis (e.g., The Northern Subject Rule). One of the three major transitory zones on the dialect map of Britain is Glauser's between Scotland and England (besides the Ribble-Humber Line splitting the linguistic north into two, and the Severn-Wash

[10] E.g., Gussenhoven and Jacobs (2011: 196) present a chart with corresponding examples from RP vs. Yorkshire vs. Durham for the purpose of an exercise in their textbook.

Line with the STRUT and BATH transition zone – see the map in Honeybone and Maguire 2020: 15). However, the case of laryngeal contact dealt with in the above discussion has not been proposed, and laryngeal phonology as such tends to be ignored altogether in the relevant literature – although, as argued above, the connection seems justified.

Analogously, then, we may expect to attest both fully-fledged voice systems (such as that of Scots) and asymmetrical, hybrid systems (such as those of Yorkshire and Durham) in other parts of the English-speaking world, too; primarily, among contact-induced varieties (ethnolects; pidgins and creoles; foreign-accented non-native Englishes) but also in (more) monolingual speech communities (as, e.g., Hunnicutt and Morris (2016) exemplifies with Southern American English). The question begs for future research whether the emergence of such intermediate, mixed or asymmetrical laryngeal settings is systematic along any kind of phonological or sociolinguistic dimension.

Acknowledgements

I am deeply indebted for their valuable comments to the anonymous reviewers of this paper as well as the audiences of various conferences between 2019 and 2022 where its preliminary versions were presented. All the remaining errors are mine. Certain phases of this research were conducted as part of the "Theoretical and Laboratory Linguistics" project at PPCU, supported by the Thematic Excellence Program of the Hungarian Ministry for Innovation and Technology (TKP2020-NKA-11); others as part of the "Laryngeal patterns in synchrony and diachrony" project sponsored by NKIF grant #142498.

References

Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh University Press.

Backley, P. 2011. *An introduction to Element Theory*. Edinburgh University Press.

Balogné Bérces, K. to appear. *Dialectal variation in English meets laryngeal typology*.

Balogné Bérces, K., and B. Huszthy. 2018. Laryngeal Relativism predicts Italian. *Yearbook of the Poznań Linguistic Meeting 4:* 153–177.

Chambers, J. K., and P. Trudgill. 1998/2004. *Dialectology* (2nd ed.). Cambridge University Press.

Clark, U. 2004. The English West Midlands: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie and C. Upton (eds.), *A handbook of varieties of English: Vol. 1. Phonology,* 134–162. Mouton de Gruyter.

Cyran, E. 2012. Cracow sandhi voicing is neither phonological nor phonetic. It is both phonological and phonetic. In E. Cyran, B. Szymanek, and H. Kardela (eds.), *Sound, structure and sense. Studies in memory of Edmund Gussmann,* 153–184. Wydawnictwo KUL.

Cyran, E. 2014. *Between phonology and phonetics: Polish voicing (Studies in Generative Grammar 118).* Mouton de Gruyter.

Cyran, E. 2017. 'Voice' languages with no [voice]? Some consequences of Laryngeal Relativism. *Acta Linguistica Academica 64*(4): 477–511.

Docherty, G., Watt, D., Llamas, C., Hall, D., and J. Nycz. 2011. Variation in Voice Onset Time along the Scottish-English border. In *Proceedings ICPhS XVII:* 591–594.

Dutta, H. and M. Kenstowicz. 2018. The phonology and phonetics of laryngeal stop contrasts in Assamese. In R. Petrosino, P. Cerrone and H. Hulst (eds.), *From sounds to structures: Beyond the veil of Maya,* 30–64. De Gruyter Mouton. https://doi.org/10.1515/9781501506734-002.

Firth, S. 1991. *Practical phonetics and phonology: Yorkshire assimilation*. [Unpublished master's thesis]. University of York.

Gussenhoven, C., and H. Jacobs. 2011. *Understanding phonology* (3rd ed.). Routledge.

Harris, J. 1994. *English sound structure*. Blackwell Publishers.

Harris, J., and G. Lindsey. 1995. The elements of phonological representation. In J. Durand and F. Katamba (eds.), *Frontiers of phonology,* 34–79. Longman.

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

Honeybone, P. 2011. *Lost in Linguistics: A guide to the current landscape of linguistic theory. Phonology*. Handouts to the minicourse held at the University of Oslo.

Honeybone, P., and Maguire, W. 2020. Introduction: What is dialect writing? Where is the north of England? In P. Honeybone and W. Maguire (eds.), *Dialect writing and the north of England,* 1–28. Edinburgh University Press.

Huber, D., and K. Balogné Bérces. 2010. [voice] and/versus [spread glottis] in the modified Leiden model. *Acta Linguistica Hungarica 57*(4): 444–457.

Hughes, A., Trudgill, P., and D. Watt. 2012. *English accents and dialects* (5th ed.). Routledge.

Hunnicutt, L., and P. A. Morris. 2016. Prevoicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics 22*(1): 214–224.

Iverson, G. K., and J. C. Salmons. 1999. Glottal spreading bias in Germanic. *Linguistische Berichte 178:* 135–151.

Iverson, G. K., and J. C. Salmons. 2003. Legacy specification in the laryngeal phonology of Dutch. *Journal of Germanic Linguistics* 15: 1–26.

Iverson, G. K., and J. C. Salmons. 2008. Germanic aspiration: Phonetic enhancement and language contact. *Sprachwissenschaft* 33: 257–278.

Jansen, W. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch* (Groningen Dissertations in Linguistics 47) [Doctoral dissertation, University of Groningen].

Johnston, P. 1997. Older Scots phonology and its regional variation. In C. Jones (ed.), *The Edinburgh history of the Scots language,* 47–111. Edinburgh University Press.

Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., and T. S. Zamuner. 2007. Representations of [voice]: Evidence from acquisition. In J. van de Weijer and E. J. van der Torre (eds.), *Voicing in Dutch. (De)voicing – phonology, phonetics, and psycholinguistics,* 41–80. Benjamins. https://doi.org/10.1075/cilt.286.03kag.

Kerswill, P. E. 1987. Levels of linguistic variation in Durham. *Journal of Linguistics* 23: 25–49.

Lisker, L., and A. S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*: 384–422.

Maguire, W. 2012. English and Scots in Scotland. In R. Hickey (ed.), *Areal features of the anglophone world,* 53–78. Mouton de Gruyter.

Masuya, Y. 1997. Voice Onset Time of the syllable-initial /p/, /t/ and /k/ followed by an accented vowel in Lowland Scottish English. In *Onseigaku to oninron: Shuyo ronko* [Phonetics and phonology: Selected papers], 139–172. Kobian Shobo.

Ouddeken, N. 2016. Voicing distinctions in the Dutch-German dialect continuum. In J. Audring and S. Lestrade (eds.), *Linguistics in the Netherlands 2016,* 106–120. Benjamins.

Piroth, H. G. 2003. Final devoicing and syllabification in German consonant clusters: A phonetic investigation. In M. J. Solé, D. Recasens and J. Romero (Eds.), *15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3-9, 2003,* 2749–2752. ICPhS Archive, http://www.internationalphoneticassociation.org/icphs/icphs2003.

Scobbie, J. M. 2005. *Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: Speech production evidence for fine-grained linguistic plasticity*. QMUC Speech Science Research Centre Working Paper WP2.

Scobbie, J. M. 2006. Flexibility in the face of incompatible English VOT systems. In L. M. Goldstein, C. Best and D. H. Whalen (eds.), *Laboratory Phonology 8: Varieties of phonological competence,* 367–392. Mouton de Gruyter.

Sonderegger, M., Stuart-Smith, J., Knowles, T., Macdonald, R., and T. Rathcke. 2020. Structured heterogeneity in Scottish stops over the twentieth century. *Language 96*(1): 94–125.

Stuart-Smith, J. 2004. Scottish English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie and C. Upton (eds.), *A handbook of varieties of English: Vol. 1. Phonology,* 47–67. Mouton de Gruyter.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., and R. Macdonald. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology 6*(3–4): 505–549.

Trudgill, P. 1999. *The dialects of England* (2nd ed.). Blackwell.

Watt, D., and J. Yurkova. 2007. Voice Onset Time and the Scottish Vowel Length Rule in Aberdeen English. In J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences,* 1521–1524. Universität des Saarlandes.

Wells, J. C. 1982. *Accents of English*. Cambridge University Press.

Whisker-Taylor, K., and L. Clark. 2019. Yorkshire Assimilation: Exploring the production and perception of a geographically restricted variable. *Journal of English Linguistics 2019*: 1–28.

Wilhelm, S. 2018. Segmental and suprasegmental change in North West Yorkshire – A new case of supralocalisation? *Corela [Online], HS-24*. https://doi.org/10.4000/corela.5203.

LingBaW *Linguistics Beyond And Within 8 (2022), 37-54*

Crossing language boundaries. The use of English in advertisements in Polish lifestyle magazines

Marta Dąbrowska
Jagiellonian University, Cracow, Poland

Abstract
Advertising is an area open to various types of innovation and a creative use of language. Advertisers use a variety of strategies to attract as many customers as possible, employing enticing visuals, clever puns, detailed descriptions, direct appeals, etc. Yet, if the product is to sell, it appears self-evident that the language of the advertisement should not constitute a barrier. Contrary to this assumption, however, there has been an ever growing trend visible in Polish magazines for the last two decades to formulate parts of or even complete advertisements, not infrequently of Polish brands, in English, despite the fact that English remains a foreign language in Poland, although, admittedly, a very popular one.

The present study is conducted within the framework of sociolinguistically informed linguistic landscape analysis. It investigates the visibility of English in the advertising landscape of lifestyle magazines published in Poland, represented by a selection of high quality and mid-range general interest and specialised women's and men's magazines. Its objective is to identify tendencies in the manifestations of this newly developing multilingualism in respect of the frequency of using English, the category of advertised products which particularly often appear in such multilingual advertisements as well as the type of magazines which admit such advertisements rather readily. Regarding the linguistic strategies used in the analysed texts, attempts are made to establish whether English tends to be used in complete texts and sentences or only parts of those, and if the latter, which elements of the genre of advertising are especially marked by such language choices and why.

Keywords: multilingual advertising, English as a global language, linguistic landscape, sociolinguistics

1. Introduction

Blommaert (2013: 44) claims that "[e]very sign tells a story about who produced it, and about who is selected to consume it." In our daily life we are surrounded by signs of multifarious and complex character. They are placed in a certain environment and their use there, deliberate or subconscious, carries a multitude of information about their past, present and future, about the character of their (functional or symbolic) use. The investigation of the "visibility and salience of languages on public and commercial signs" (Landry and Bourhis 1997: 23), which initiated a new field of study called linguistic landscape (LL) towards the end of the previous century has grown into a rich and complex area of research. It is currently

more aptly defined as "attempts to understand the motives, uses, ideologies, language varieties and contestations of multiple forms of 'languages' as they are displayed in public spaces" (*The Linguistic Landscape: An International Journal* 2015).

The LL framework appears to be a particularly useful tool to investigate how developments of globalisation have been shaping the use of local languages in the face of the influence of the English language, the most important *lingua franca* of the contemporary world. Its growing presence and prestige in countries which were not a natural area of its use a few decades ago add up to the linguistic repertoire of the local users. As a result, English has been affecting their language choices and uses in various domains and genres, especially those not so strictly regulated by official linguistic norms of a given country (cf. Sebba 2012), thereby also influencing orders of indexicality[1] of the respective languages in various fields (cf. Blommaert 2010). As regards Poland, advertising appears to be an area in which, due to the assumed flexibility and linguistic creativity marking the genre (cf. Sebba 2012), introducing other languages, notably English, beside Polish, seems to have been particularly easy and relatively unchallenged, despite the existence of an official language act.[2] The following analysis aims to present a preliminary investigation of the linguistic landscape of advertisements found in Polish lifestyle magazines. It will include top-shelf and mid-range magazines of general and more specialised interest, addressed to both women and men. The objective of the study is to analyse the content of the advertisements, treated here as examples of multilingual texts (cf. Sebba 2012), in terms of language choices found in various elements of the genre, and investigate the character (i.e. the frequency, form, and function) of English items found in the texts. A sociolinguistic analysis of the above elements will also make it possible to study the distribution of English and Polish in respect of the type of goods and services advertised, and the quality of the magazines in which English appears in the advertisements most often. This, in turn, will indirectly allow us to assess who the addressees of the texts composed partly or fully in English tend to be in terms of their possible social status.

---

[1] The concept of *indexicality* has been proposed by Silverstein (2003: 195), who argues that "any linguistic, a.k.a. sociolinguistic, fact is necessarily an indexical fact, that is, a way in which linguistic (…) signs-in-use point to contexts of occurrence structured for sign-users in one or another sort of way." Orders of indexicality are "foci of normativity" in polycentric sociolinguistic systems (Blommaert 2013: 11).

[2] Article 7 of the Act on the Polish language of 7 October, 1999 stipulates that "[o]n the territory of the Polish Republic, the Polish language shall be used in all legal action between Polish entities as well as in cases when one of the parties is Polish. Specifically, this paragraph refers to names of goods and services, advertisements, instructions for use, information about properties of goods and services and warranty terms and conditions, invoices, bills and receipts." In 2014 a viewer appealed to the court with a complaint about the fact that the actors in a commercial of Paco Rabanne perfumes spoke English on Polish TV, while, following the Act on the Polish language, advertising should be provided in the language understood by everybody in the country and on this ground he won the case (*Rzeczpospolita*, 6 November, 2014).

2. Studying the use of English in advertising

A tendency to use English in advertising next to local languages is naturally not limited to the Polish context alone. Similar observations have been made both in the context of multilingual countries (e.g. Grin 1994, Ye and Qin 2004) and other non-English speaking countries affected by the processes of globalisation (cf. Cheshire and Moser 1994, Gerritsen et al. 2007, Haarmann 1989, Kelly-Holmes 1998, 2000, Piller 2001, Ustinova and Bhatia 2005, Zhang 2001). When analysing the use of English in Russian advertisements, for instance, Ustinova and Bhatia (2005: 496) observed that some elements which were primarily kept in English were company names, product names, and labelling. In their view, those phenomena were to be seen as manifestations of the processes of globalisation vs. customisation (or standardisation vs. adaptation) when advertising foreign brands (cf. Grant and Short 2002, de Mooij 1994). In the former, brands keep the original form and content of the advertisement when entering foreign markets (cf. Kotler et al. 2002), while the latter shows some local modifications of the text (cf. Baumgardner 2008). The use of English in advertising can also be viewed in terms of two similar parallel processes: homogenisation, which is especially linked with spreading and adopting the American model of brands in the sense of accepting a certain vision of the world, and hybridisation, in which English and the cultural model it indexes are mixed with the local models and languages (Bhatia and Ritchie 2004, Lee 2000). In his overview of the use of languages in advertising Spolsky (2009: 35) argues that it would only be logical to expect that "the normal result of a communicative goal would be a policy to advertise in the language of potential clients and customers," as a thorough comprehension of the message would naturally boost the sales of the advertised products (cf. Grin 1994). However, a number of studies of the language of advertising or public signs in shops and services show that foreign languages, and notably English are used there not for communicative function, but because of their symbolic or indexical value, as such a choice often connotes modernity, high quality and glamour (cf. Bhatia and Ritchie 2004, Dąbrowska 2020, Haarmann 1989, Kelly-Holmes 2000, 2005). In fact, as Gerritsen et al. (2000) point out, understanding the message of the advertisement is not of primary importance for marketers, since they are more interested in creating brand awareness and positive attitude to the advertised product, rather than in a full comprehension of the text.

Overall, the main focus of studying the use of foreign languages, and primarily English, in advertising has been the symbolic use of the language (Haarmann 1989), but also, more currently, the process of identity construction of the addressees (Piller 2001, 2003) as well as aspects of globalisation (Bhatia and Ritchie 2004, Blommaert 2010). An important observation stemming from the above is that what seems to inform the modern linguistic developments worldwide is a need to search for social identity and self-definition. Studies of advertising also show that various items which are advertised are linked with certain values, a lifestyle that users seek (cf. Vestergaard and Schroder 1985). The use of English marked by its high present-day indexicality associated with such consumer goods (Blommaert 2010) seems to contribute to meeting this need significantly.

A study which highlighted those trends already three decades ago, and which has also inspired the current investigation was one conducted by Cheshire and Moser in Switzerland

(1994). The researchers focused on the analysis of advertisements in two Swiss magazines, one more oriented to politics and the other to sport, lifestyle, etc. What they discovered was the fact that 7% of the advertisements were drafted in English alone, 20% had only the name of the product in English, and 73% demonstrated a mixture of both English and one of the local languages, usually French. In the latter category English was quite prominent, primarily in phrases that expressed slogans of the advertisements. In the researchers' interpretation, the use of English they investigated was different from that in Haarmann's (1989) study of the use of English and other languages in Japan, which he had identified as a cultural symbol. Here it was viewed as a "language display" (cf. Eastman and Stein 1993) "laying claims to the attributes associated symbolically with speakers of English" (cf. Garrett 2010: 145). Their analysis of the study material demonstrated that English was employed more often to advertise transient fashions (ibid.), e.g. clothes or cars, than everyday products, etc., it was also linked with travel and tourism (e.g. credit cards were always advertised in English) as well as with technological developments (communication, computers, hi-fi equipment, etc.). Some advertisements carried clear connotations with English-speaking countries (e.g. Scottish whisky), however, what is especially significant in the context of the present study, some Swiss companies advertised their products in English as well (32% of them did it in English to a French-speaking readership, including advertisements of Swiss watches).

Investigating the visibility of English in Polish advertising has been gradually attracting some research interest in the last twenty years too, although, as Planken et al. (2010) point out, there has been relatively little research on the topic in Eastern European countries overall (cf. Griffin (2001) for Bulgaria, Kelly-Holmes (2005) for the Czech Republic, Ustinova (2006, 2008), Ustinova and Bhatia (2005) for Russia). An early study of the use of English in Polish advertisements was undertaken by Griffin (1997).[3] It showed that 88% of the 346 analysed advertisements contained at least one word in English. Similarly, in her MA thesis Bulawka (2006) demonstrated that 90% of 235 advertisements that she analysed contained some element of a foreign language, and in 79% of the advertisements such elements were in English. This only left 10% of the advertisements entirely in Polish, and they were those which mainly advertised food products. As Bulawka (2006: 19) established, the items which were expressed in English, or in a Polonised form of an English word, ranged from single items (mostly names of brands, logos, slogans, elements in the main body of the text, the header, the subheader), through more elaborate phrases to entire texts.

On the other hand, in her MA thesis Tobolova (2012) undertook to analyse the use of English in Polish TV commercials, and in particular, in commercials of cosmetics, food, electronic equipment, cars, and pharmaceuticals of both foreign and Polish origin. The investigation demonstrated that the highest ratio of the use of English (detected in 100% of the commercials) was found in advertisements of cosmetics. Further, English was utilised in 94%

[3] *Angielski w polskiej reklamie*, a well-known book by Chłopicki and Świątek (2000), will not be included in the overview, since, as the authors themselves explain (2000: 16), it is not a book directed to linguists or those interested in a linguistic analysis of advertisements. The main goal of their publication is to identify and rectify typical grammatical, lexical, semantic and other mistakes found in the English elements used in the Polish advertisements analysed in the book. The prescriptive and pedagogical angle employed in the publication does not match the descriptive sociolinguistic approach followed here.

of advertisements of electronics, 78% advertisements of cars, in 54% of food advertisements, and only in 10% of pharmaceuticals. The commercials were mainly of foreign products, and, apart from a minor informative function they carried, they were found to predominantly enhance the attractiveness of the product. In the case of cars and electronics there was a high frequency of English slogans observed, while in food commercials English was used mainly in accompanying songs. There were few Anglicisms used in the case of pharmaceuticals, because, as the author concluded, those commercials were directed to senior viewers, who seldom speak English.

Planken, van Meurs and Radlińska (2010), in turn, analysed the effect which the use of English in Polish advertisements had on the readers. The authors wished to investigate whether claims that the use of English in advertising would enhance the selling value of the product would be corroborated by their study results. In order to achieve their aim they researched attitudes to a given product/brand image and to the advertisement itself; they also evaluated the respondents' purchasing intentions, as well as the degree of their text comprehension. The analysis was conducted with regard to six sets of advertisements drafted in English, found in glossy magazines, and their parallel versions translated into Polish. The data were collected by means of a questionnaire distributed among 62 Polish female students of Polish studies. The results obtained did not, in fact, demonstrate any significant differences with regard to the investigated variables in respect of the language of the advertisements. The initial assumption that the use of English in advertisements had positive effects on non-English speaking target groups as opposed to their native language was, therefore, not upheld. The researchers point out, however, that such results may have been an outcome of the high homogeneity of the investigated group (young highly educated women), and the effect may have been different, had the investigation been conducted on a more diverse population.

The present study recognises the contribution of the investigation of the use of English in Polish product advertisements conducted so far. A justification to take it forward is to be seen in the fact that the frequency of the use of English in Polish public space has been growing, leaving its mark in various elements of linguistic landscape (cf. Dąbrowska 2020), and a duty of a sociolinguist is to observe the changing tendencies in the linguistic environment as they happen. With the widespread use of English it is now possible to investigate its visibility also in respect of who the recipients of such texts are in terms of their social status. The current study will attempt to establish it as well, beside identifying and discussing the frequency and character of English items found in the analysed texts. This will be achieved through an overview of which type of magazines tend to admit multilingual advertisements more readily than others.

At this stage it is necessary to clarify the analytical approach which this study will follow. So far, as Androutsopoulos (2012) points out, most studies of the use of Anglicisms[4] have

[4] The concept of *Anglicism* is understood here as an English word incorporated into another language and used as a part of its lexicon. This understanding is supported by Gottlieb (2005: 163), who postulates that Anglicism is "any individual or systemic language feature adapted or adopted from English, or inspired or boosted by English models, used in intralingual communication in a language other than English." In this sense it is to be viewed as a synonym of an English borrowing (of either the morphological, semantic or syntactic type).

been focused on lexical aspects, and English phrases or clauses have not been a part of research. He considers this development regrettable in view of the fact that an adequate interpretation of multilingual texts, especially the genre of advertisements of various kind, also involves multimodal aspects, for instance the location of various items used in the analysed data, their function there, their size, colour and other aspects that vary in respect of the languages utilised there. Therefore, the focus of such studies should not be on borrowings, but on strategic language choices found in respective elements of the analysed texts. This is also an approach employed in the analysis presented here. Consequently, the study does not investigate the material from the point of view of the theory of borrowings, and hence no literature concerning borrowings is quoted in support. A view which informs this study is that borrowings are treated as elements belonging to the borrowing language, they are used in keeping with the norms of the borrowing language grammar and understood by monolingual speakers (cf. Hudson 1996, Poplack 1988, Sankoff 2001). Texts including borrowings are, therefore, treated as monolingual too, whereas the advertisements studied here are treated as multilingual[5] (cf. Sebba 2012), that is why comments will be made about languages utilised in the texts in their various parts and functions, not about borrowings. A focus on borrowings might be entailed by an investigation of the phenomenon of code-switching, though the character of the relationship between the two has not been agreed on (cf. Gardner-Chloros 2009). However, while references made to the phenomenon of code-switching[6] might appear applicable in the present context, they will be adduced only briefly in the investigation of examples whose internal structure might be analysed in respect of intra-sentential code-switching (cf. Myers-Scotton 2009, Poplack 1980), or code-mixing (cf. Muysken 2000). According to Sebba (2012), a study of written multilingual texts is not to be viewed as an analysis of code-switching, either (its investigation may only constitute an element of the overall complex structure analysis), because of the additional multimodal aspects underlying a

Borrowing, in turn, according to the classical definition offered by Haugen (1950: 212), is to be understood as "the attempted reproduction in one language of patterns previously found in another."

[5] Naturally, the texts analysed here are only bilingual, in Polish and English. The use of the terms *bilingual* and *multilingual* varies depending on the author, with some using the former as a general reference to the use of more than one language, and others the latter (cf. Mesthrie 2000, Trudgill 2003). In the present study it is the term *multilingual* that is used as an umbrella term for any linguistic situation which involves the use of two or more languages. Such an approach is found to be inclusive of texts of advertisements which are composed in more than two languages. A use of more than two languages in an advertisement is a frequent situation, especially in multilingual countries (cf. Sebba 2012), but is also gradually becoming the case in Poland, where, beside Polish and (frequently) English, also Ukrainian starts to be incorporated in the same advertisement, especially in outdoor advertising.

[6] The phenomenon of *code-switching* and its structural and functional categories are a challenging issue to analyse, which is reflected in a complex terminology describing it, postulated by various linguists representing different theoretical backgrounds (cf. Gardner-Chloros 2009). As the phenomenon will not be a primary analytical focus of this study, a simple definition which describes it as an "alternation between the different varieties which people speak" should suffice (ibid.: 11). A distinction between *code-switching* and *code-mixing*, which is observed by some authors, and the latter is also used in this study, may be broadly described as a distinction between the use of two languages in speech when the two codes maintain their monolingual characteristics, and a situation in which the two systems show some convergence within an utterance (cf. Muysken 2000).

proper understanding of such texts, already highlighted above (cf. Androutsopoulos 2012). It should rather constitute an analysis of language choices and their distribution in a multilingual and multimodal linguistic landscape in which they were located.

3. Methodology and data

The present analysis is based on an overview of 21 issues of altogether 14 Polish high quality and mid-range[7] general interest and more specialised lifestyle magazines: *Twój Styl* (2 issues), *Vogue*, *Wysokie Obcasy, Viva* (4 issues), *Świat Kobiety, Pani*, *Przyjaciółka*, *Tina*, *Rewia* (2 issues), *Men's Health* (3 issues), *KiF*, *Motocykl*, *Dobre Wnętrze, Moje Mieszkanie*, (all of the issues were collected between June and October 2021).[8] The principle behind the above selection was not to provide a thorough and exhaustive overview of all the titles and issues available on the market. For the purpose of this preliminary type of study they were rather to represent a cross-section of the potential readership they are addressed to in terms of their gender (women, men), economic aspects (top quality vs. mid-range) as well as a variety of topics the readers are interested in (lifestyle, cosmetics, fashion, health, sport, entertainment, interior decoration, etc.). Such a selection, in turn, was to allow for an overview of the language choice tendencies depending on the target audience and topic of the respective magazines, as already pointed above. This is an aspect of investigation not considered in respect of studying language choice in advertising before (the category of the addressees and the more specialised topics of the magazines discussed here were not variables in the previous studies of English found in Polish advertisements).[9] The selected magazines also represent a

[7] *Merriam-Webster Dictionary* online defines quality magazine as "a periodical containing material designed to appeal especially to readers of superior education or culture," while, according to *Collins English Dictionary* online mid-range products are those which are "neither the most expensive nor the cheapest of their type." In the present study the main defining element distinguishing between the two categories of journals selected for analysis was their price, ranging from magazines like *Vogue* (19.99 PLN), *Wysokie Obcasy* (9.96 PLN), *Twój Styl* (9.49 PLN), and *Men's Health* (7.99 PLN), to *Viva* (4.99 PLN), *Świat Kobiety* (3.49 PLN), *Tina* (2.69 PLN) or *Rewia* (2.19 PLN). Consequently, the break-off point between a high quality vs. a mid-range magazine was, respectively, its price above or below ca. 8.00 PLN. Additionally, the classification of the respective magazines to one of the two categories was based on their number of pages as well as the content of the analysed periodicals. Beside texts about fashion and cosmetics, quality magazines also feature editorials, specialised columns, essays on culture, and interviews with inspiring people, while mid-range periodicals contain, primarily, cooking recipes, fashion ideas, household maintenance advice, health topics, crosswords, and gossip about celebrities.

[8] The selection of the magazines was random, provided they fitted in the broad category of lifestyle magazines, and were available for sale during the period of the study. The more bulky magazines were primarily represented by one issue (except for *Twój Styl*, which came in two different sizes), smaller ones needed a greater representation because of the low number of pages, similarly as the magazines addressed to men, due to a much more limited number of men-oriented titles available on the market.

[9] Previous studies of English elements in Polish advertisements to date primarily investigated only the texts which were known to contain English elements. The starting point for the present analysis is an overview of all the advertisements in the magazines selected for analysis, first irrespective of their language, but focusing on the quality of the magazines and topics tackled (general vs. specialised) first. This, in turn, will allow us not

variety of options in terms of their quality, size, frequency of publication as well as foreign vs. local ownership (viz. *Vogue* vs. *Wysokie Obcasy*, *Men's Health* vs. *Moje Mieszkanie,* or *Twój Styl* vs. *Rewia*). In this way the study material also takes into consideration potential differences in the social standing of the target readership of the magazines. This, in turn, influences the selection of the advertised products in terms of their necessity and price, as well as the possible choice of the language used to advertise the products too. The following analysis will first focus on the visibility of English in the collected samples of the advertisements in terms of its statistics in respective magazines. Subsequently, the types of products in advertisements of which English was employed will be identified. Finally, the analysis will take into consideration the share of the use of English (from one item to a full text) as well as the form and function of the English items found in the texts.

4. Analysis of data

The overall number of advertisements found in all the magazines is 348. When broken down into particular magazines, the distribution of the advertisements is as follows (from the highest number to the lowest in total, with a calculation per issue where applicable): *Twój Styl* – 83 (ca. 42 per issue), *Świat Kobiety* – 32, *Vogue* – 31, *Pani* – 25, *Wysokie Obcasy* – 23, *Przyjaciółka* – 13, *Men's Health* – 39 (ca. 13 per issue), *Viva* – 49 (ca. 12 per issue),*Tina* – 10, *Motocykl* – 10, *KiF* – 9, *Dobre Wnętrze* – 9, *Rewia* – 10 (5 per issue), *Moje Mieszkanie* – 5. The above distribution shows immediately that brands typically choose magazines of higher quality in order to advertise their products, including those of a more luxurious character; the more specialist magazines tend to have fewer and more thematically related product advertisements. While *Rewia* and *Moje Mieszkanie* have the lowest number of products advertised, including mainly medication supplements and food products or else elements of interior design (floors, curtains, etc.), respectively, the other ones mainly specialise in luxury items, like cosmetics, clothes, watches, jewellery, cars, hotels and spas (see below).

The overview of the language used in the analysed texts demonstrates a frequent use of English, with 194 out of 348 (55.74%) advertisements containing at least one English element. The list presented below shows how many of all the advertisements found in the respective magazines were marked by the use of two languages or of English alone, with English ranging from one item to the whole text composed in it. The ranking presents itself as follows: *Dobre Wnętrze* – 8/9 (88.88%), *Vogue* – 26/31 (83.87%; fully in English – 7/31 – 22.58%), *KiF* – 7/9 (77.77%), *Men's Health* – 29/39 (74.35%), *Pani* – 18/25 (72%), *Motocykl* – 7/10 (70%), *Twój Styl* – 57/83 (68.67%; fully in English 6/83 – 7.22%); *Moje Mieszkanie* – 3/5 (60%), *Viva* – 26/49 (53.06%); *Wysokie Obcasy* – 11/23 (47.82%); *Tina* – 1/10 (10%); *Przyjaciółka* – 1/13 (7%), *Świat Kobiety* – 0/32 (0%); *Rewia* – 0/10 (0%). It is thus interesting to observe that the magazines at the top of the list, with the highest share of English, are predominantly those more focused on specialized fields. They contain a low number of advertisements, yet their specialized character, and possibly a higher share of technical details they contain, clearly

only to confirm that English is used in Polish magazines, but also establish in what types of periodicals it appears most, what products it tends to be associated with, and what type of readers it typically targets.

entail the use of certain concrete ingredients or possession of certain features known globally under their English names; alternately, they may be just names of brands. Examples in this category typically constituted individual words or short phrases (e.g. *Rillington/Wembley/ Caramel* (types of oak)*; plug-in hybrid; ice protein shake; www.balticus-watches.com*;[10] *Melitta Barista TS Smart*), although occasional longer chunks of English texts were found too. A more interesting situation was to be observed in advertisements placed in quality magazines, notably in *Vogue*, *Twój Styl* and *Pani*, in which English appeared in clearly more than 50% of the texts, and a significant use of all-English advertisements could be found there as well (though, admittedly, short ones, as will be elaborated on below). At this point it is worth pointing out the difference in the share of advertisements in, e.g. *Wysokie Obcasy* (viz. 47.82%), which is owned by a Polish publisher, and e.g. in *Pani*, owned by a foreign capital, with 72% of such advertisements. The visibility of English in about 50% of the advertisements placed in mid-range magazines, like *Viva* is notable too. On the other hand, the fact that the inexpensive magazines showed close to none or no visibility of English whatsoever is also very telling. It demonstrates that luxury brand owners make conscious decisions as to what kind of readership they want to target and what strategies they need to use when trying to sell their products. It may be surmised that it is not only that they advertise their products in magazines bought by more well-off customers, but they also aim for more educated buyers, since such persons are more likely to understand foreign texts or those with some foreign colouring added to them.

An overview of the advertisements which show some visibility of English, either partly, or fully, has demonstrated the following distribution of the products in the case of which marketers have considered it justified to advertise them with the help of English. The list, in terms of the number of advertisements promoting respective products identified in the researched magazines, presents itself as follows: cosmetics – 53, clothes – 25, motorcycles – 10, shoes – 9, perfumes – 9, watches – 9, cars – 7, events – 7, furniture – 4, water – 4, accessories – 3. The remaining categories: jewellery, coffee, lingerie, aesthetic medicine, bathrooms, kitchen, hotels, films, radio stations, and food were advertised with the use of English once or twice each. Certainly, the above distribution confirms to a large extent that English is used to advertise products which are luxurious in character and do not just meet people's basic needs (cf. Cheshire and Moser 1994). Interestingly, though, the selection of goods advertised in the Polish magazines as opposed to those that Cheshire and Moser (ibid.) studied is overall quite different (e.g. no credit cards, alcohol, cigarettes, etc., have been found in the current study, which no doubt also shows that the fashions and needs are changing with times).

Table 1 below summarises the results of the analysis of the collected advertisements in respect of the visibility of English and Polish in their content:

[10] Although this item is a website address, it contains the name of a (Polish) brand Balticus Watches, which is naturally an English phrase. Due to the fact that it constitutes an element of the advertisement body copy, is highly visible due to the size of the font, and may potentially take the reader to the company website, it has been included among the English items analysed here as well.

Table 1: *Distribution of English items in the analysed advertisements*

Degree of English visibility	**Number of advertisements**
A total number of advertisements with at least one English item:	194 (55.74%)
1. English only advertisements:	37 (10.63%)
A. Only the brand/product name and an accompanying an image	23 (6.60%)
B. A complete text in English	14 (4.02%)
2. Mixed-language advertisements:	157 (45.11%)
C. One English item in a Polish text	96 (27.58%)
D. Several items in a Polish text	61 (17.52%)

Thus, as regards the degree of visibility of English in the studied texts, the above overview has demonstrated the following share of the use of the two languages: there were 156 advertisements fully in Polish and they constituted only 44.82% of the advertisements scanned. On the other hand, as indicated above, 194 advertisements, i.e. 55.74% of the investigated material, contained at least one English item. The latter group can, in turn, be subdivided into two main categories: (1) English only advertisements and (2) mixed-language advertisements, each of which is further broken down into two subcategories. The first subcategory of English only advertisements is a group of texts in which the image is accompanied by only the name of the brand or its product, and the name is an English word (there were 23 such items, which constituted 6.60% of all the texts). The other subcategory of English only advertisements contains those in which the whole text was written in English (14 items – 4.02%). As far as the mixed-language advertisements are concerned (which, in fact, constitute the majority of the analysed texts), as said above, they also consist of two subcategories. The first one, with 96 advertisements (27.58%), contains one English word in otherwise a Polish text, and the other one, consisting of 61 advertisements (17.52%), makes use of a few items/phrases embedded in a Polish text.

Expectedly, the number of advertisements only in English is fairly low – 37 in all (10.63%), with the above-indicated division between the advertisements which consist of (A) only the name of the product/brand and (B) longer texts written only in English. Examples of the first category, repeated in various magazines, include: *Guess, Marccain*, *Solar, Pinko, Blauer, MaxMara,* etc., though indeed, sometimes it is hard to decide what language a given name can be derived from. The fact that some brands (e.g. *Solar* in the analysed sample) are Polish companies is also indicative of the popularity of English in Polish fashion industry and may lead to a conclusion that such companies attempt to get recognition on the global market without indicating their actual origin straight away. The latter category (i.e., longer texts written in English) may be illustrated by advertisements of such brands as, e.g., *Gerry Weber, Jenny Fairy, Blauer USA, New Balance, Longines, Ecco* or *CCC*. An example (from *Ecco*) reads as follows:

> *Ecco Modtray* (in a bubble)
> *Slow fashion. Just in*,
> *ecco.#MovesLikeYou*
> *ecco.com*.

Another illustration comes from Gerry Weber:

> *Gerry Weber. I live my style.*
> (partly on the image): *We are Gerry*
> *gerryweber.com*

Indeed, the number of English-only advertisements is not overwhelming, and must have been motivated by the globalisation trend to be found in the foreign brands which have used the same advertisements in various countries. It needs to be remembered, however, that the advertisements are presented in a foreign language to Polish recipients in a country which is officially monolingual, and whose law stipulates the use of Polish in public contexts. Such a choice on the part of advertisers is, then, indicative of an implicit assumption that English in globally understood. It is also a significant fact that such advertisements were predominantly found in luxury magazines: *Vogue* and *Twój Styl*. This observation, in turn, indirectly shows that English tends to be used only with a fraction of Polish society, i.e. with the educated and well-off social group.

As indicated above, the most prominent category which the analysed advertisements fall into is a mixed-language group, i.e. one which contains texts in Polish with elements of English. Altogether, such texts were very evenly distributed among all the magazines, and comprised 157 such advertisements, i.e. 45.11% of all the collected material. These, as already mentioned above, need to be further subcategorised into (C) mixed texts in which the English elements are only the names of the advertised product(s) or the company that produces/offers them, i.e. 96 texts (27.58%), and (D) those which, beside names of brands/products, contain another item in English, e.g. the slogan, the signature, or some additional product information (61 items, i.e. 17.52% of all the collected advertisements). The latter is not a very high ratio either, yet, when one realises that such details may not be fully understood by the readers, i.e. potential buyers/users of the products, the figure is to be viewed as significant. Examples of the category where only the name of the product is foreign include: *RedBlocker, ArtiShoq, GoodValley, Sport Camp, UltraBlanc, Be Bio, Natural Nation*, *Focus*, *Endurance*, etc. Examples of brands which, in turn, apply English in the names of their products are illustrated by the following list, along with the respective product names provided in brackets: AA (*Enjoy Nature*), Ava (*Botanical HiTech*), Bielenda (*Good Skin*), Dr Irena Eris (*Institute Solutions*), Balticus (*Balticus Watches*), Dermika (*Insomnia*), Paradyż (*My Way*), Ziaja (*Ziaja Baltic Home Spa. Wellness*). Moreover, as regards the mixed-language category, it is important to observe that most of the companies/products listed here are Polish, which demonstrates that their choice of the company or the product name in English undoubtedly aims to heighten the prestige of the brand through associations with its international status, and a better quality of their product implicated by it. It possibly also gives them a greater chance of recognition and circulation of their products abroad. It is important to note that, altogether, 17 Polish brands have chosen to use English either in their names or in the advertisements of their products. Some examples of such texts include:

- Natural Nation:
  Natural Nation.
  Focus.
  Poprawia pamięć roboczą
  (szałwia lawendolistna)
  Poprawia koncentrację umysłu
  (bakopa drobnolistna)
  Wspiera równowagę psychiczną
  (miłorząb dwuklapowy)
  20% rabatu
  Lepsza pamięć i koncentracja
  Wejdź na naturalnation.pl, wpisz kod MHprint8 I odbierz 205 rabatu. Ważny do 15 sierpnia.

- Ava:
  Ava Laboratorium,[11] *Since 1961,*
  (an image of the product with the name in English below, followed by further details): *Botanical HiTech.*
  Technologie przyszłości,
  Fuzja natury i nauki. Skuteczna redukcja zmarszczek. Innowacyjne rozwiązania kosmetologiczne.
  Dostępne w wybranych drogeriach na terenie całego kraju i w HEBE.
  We współpracy z jednostkami naukowo-badawczymi.

As regards the texts with a mixed-language message, it is hard to identify one pattern according to which the advertisement tends to be composed, based on an overview of the sample texts. Occasionally (viz. in 6.89% of the advertisements), beside the name there tends to be a slogan either following the name or added in a different font and format in the background; it may also be written in a smaller font below the body copy, often with a hashtag. The following examples illustrate the category:

- BlackRedWhite:
  (in the right upper corner the logo of *BlackRedWhite*)
  (on top of the page): *Fashion Trends*
  (below the image): *Mój styl. Moje wnętrze. Poznaj limitowaną kolekcję mebli wypoczynkowych. Zainspiruj się.*

- Miss Dior:
  Miss Dior. Nowy zapach.
  #WakeUpForLove (then the logo of Dior below the image of the perfume).

- Esprit:
  Esprit.
  Wherever you go.
  (the price and the logo of Deichman and at the very bottom another slogan in Polish): *Bo kochamy buty.*

- Sephora:
  Odkryj nowości sezonu.
  Sephora. The Unlimited Power of Beauty
  sephora.pl

[11] In this phrase the order of the items used in the name reflects the word order of nominal adjuncts typical of English. Moreover, the name is followed by the information about the date of the company's establishment, which is also provided in English.

It is in this subcategory of mixed-language advertisements that instances of code-mixing[12] can be found. They involve not only the name of the product in English, but also references to some of its features/ingredients in otherwise a Polish matrix sentence. The following examples come from Giorgio Armani, Neonail, Bielenda, and Hero:

- Giorgio Armani:
  Giorgio Armani.
  (below the image of the product): *Luminous silk. Mistrzowski podkład rozświetlający. Największy sekret Makeup Artystów. Legendarny efekt Armani glow.*

- Neonail:
  Neonail.
  Nowość. Naturalnie wegańskie! Lakiery klasyczne.
  77% składników pochodzenia naturalnego na bazie pszenicy, manioku, trzciny cukrowej oraz kukurydzy.
  Kissed by nature
  Dowiedz się więcej na neonail.pl. Lakiery plant based wonder od NEONAIL znajdziesz w centrach handlowych, perfumeriach Douglas oraz w najlepszych drogeriach w całej Polsce i na neonail.pl.
  (Additionally, in the top right area there are certain symbols in bubbles with the following texts):
  cruelty free.
  Vegan.
  77% składników pochodzenia naturalnego.
  30 kolorów prosto z natury.

- Bielenda (a Polish brand):
  Bielenda.
  Good skin
  Wypróbuj nasz 7 day miracle skin program i poznaj efekty – WOW! Już po tygodniu stosowania zauważysz: poprawę nawilżenia skóry o 24%; poprawę jędrności skóry o 14%; poprawę elastyczności skóry o 12%; wyrównanie kolorytu skóry o 3%.
  Szukaj w drogeriach Hebe.
  Skuteczność potwierdzona w programie badań aparaturowych. Efekty widoczne przy stosowaniu kremu, toniku i serum linii Hydra Boost przez 7 dni.

- Hero (a Polish brand):
  Hero.
  www.cannabishero.pl.
  Be a hero. Join team Hero.
  Hero to pasja i doświadczenie, wsparcie tych, którzy stawiają sobie wyzwania.
  Zobacz nasze supermoce.
  533339694. sklep@dobrekonopie.pl.

[12] As highlighted earlier, a proper investigation of multilingual texts, especially of advertisements, involves not only an analysis of textual aspects, but also multimodal ones (cf. Sebba 2012). In view of this a reference to code-mixing is made only in the case of examples in whose text both Polish and English are used within a sentence and/or in the complete body copy. Items like *#WakeUpForLove* or *The Unlimited Power of Beauty*, on the other hand, are instances of slogans, which not only constitute separate, self-contained sentences, but are also marked off from the remaining part of the text by a different font, its different size or colour, and often a different spatial arrangement in the advertisement, characteristic of the genre. For this reason, they are not treated as examples of code-switching or code-mixing in this study.

Occasionally, the text in the advertisement may be divided rather clearly into an English and a Polish part (cf. Sebba's (2012) category of disjoint advertisements), as in the advertisements of Certina and New Balance.

- Certina:
  When taking responsibility: Count on me.
  Supporting partner Sea Turtle Conservancy.
  Certina wykorzystuje wizerunek żółwia – symbolu wytrzymałości – od 1959 roku, od czasu wprowadzenia słynnego system zabezpieczeń DS Concept. Jako partner STC – wspiera organizację zajmującą się badaniem i ochroną żółwi morskich. CERTINA.PL.
  DS. Action Diver – Special edition, automatic movement, anti-magnetic nivachron technology, diver's watch 300 m, ISO6425, Swiss made.
  Certina. Swiss watches since 1888.

- New Balance:
  Never run out of comfort.
  Fresh foam. 1080.
  New Balance (plus a logo).
  Dostępne na www.nbsklep.pl oraz w wybranych salonach New Balance.

Based on the analysed advertisements of the mixed-language category it may be concluded that the elements in English tend to be, in the first place, names of the company and/or of the product, slogans, and often the more technical or specialist specification details. They sometimes appear as elements of the image placed in the advertisement which shows the product itself, so the boundary between the actual image of the product and the text describing it is often hard to draw (e.g. Trec Nutrition products like *Ice Protein Shake* or *Boogieman Shot*, Dermika's *Insomnia* or Ava's *Botanical HiTech*). The texts analysis demonstrates that the embedded English elements typically retained their grammatically correct structures, and so did their Polish co-text, too. The only rather unusual grammatical form found was the one in Giorgio Armani's perfume description, in the phrase *Największy sekret Makeup Artystów*. In the analysed phrase the word *makeup*, although used as a modifier of a Polish head noun, does not appear in an adjectival form, which it should have in the given position (otherwise it should follow the head noun and obtain an inflectional ending, viz. *artystów makeupu*), but it retains its nominal adjunct function which it would have in English. This makes the structure stand out as rather odd and unique.

5. Discussion

There is a considerable visibility of English in the advertisements placed in Polish life style magazines. 55.74% of the analysed sample of the 348 advertisements demonstrated the use of at least a single word in it. The decision to publish advertisements with English elements seems to depend, however, on the prestige of the magazines. Inexpensive how-to and gossip magazines addressed to readers with a lower income, as e.g. *Rewia* or *Tina* hardly feature texts with elements in English. This, in turn, is linked with the category of products advertised there, and possibly, indirectly, with the social standing of potential readers. The quality

magazines, on the other hand, notably those addressed to women (112 samples, 32.18%), less typically to men (43 samples, 12.35%), are marked by a high share of such advertisements.

The use of English ranges from individual names of brands/products, either found in advertisements once or repeated throughout otherwise a Polish text, through longer phrases, particularly slogans (24 items, 6.89%) or, much less frequently, endorsements (3 items, 0.68%), to almost full sections of the text composed in English side by side with Polish ones. As indicated above, in a limited number of cases the whole text of the advertisement is written in English (14 texts, i.e. 14.02%).

An important observation to be made is that numerous Polish companies (viz. 17 in the collected material) advertise their products with the use of English. It may either be reflected in the name the company has chosen for itself or, more frequently, for the products (their images added to the text, bearing names and information in English, enhance the brand's visibility and its impact on the potentially international market). It is also common for the advertisers to make use of words or phrases (names, slogans, names of ingredients, etc.) in English. In such cases the English items typically retain their original form and do not get inflected or Polonised. They tend to take a form of nominal adjuncts, and at times of very brief and syntactically simple sentences, most often in the imperative form (e.g. *The unlimited power of beauty; kissed by nature; good skin; cruelty free; be a Hero. Join team Hero; Never run out of comfort*, etc.). Occasionally (viz. in 4.59% of the samples), longer chunks of texts (a sequence of clauses or sentences) are written in English too. The number of relevant examples with such features in the analysed sample was not high enough to allow us to draw any reliable observations as to which elements tend to be used in English in such cases. However, a possible conclusion might be that the description of technical details in the body copy and typically the slogan may appear in English, much as in the original advertisements of foreign brands, while practical information concerning the form of distribution of the given product is predominantly phrased in Polish to allow potential buyers to locate the items easily.

What is also significant is that the messages conveyed by the two languages in the analysed advertisements on the whole differ, rather than repeat the same information, thereby complementing each other (cf. Sebba's (2012) disjoint texts). It may, therefore, be concluded that advertisers assume that their potential buyers/consumers will understand the overall message provided in the two languages without much difficulty. Since it is luxury products that tend to be advertised by means of multilingual texts, it may be reiterated that this policy indirectly indicates that advertisements of such products are first aimed at educated, and thereby also possibly better-off consumers. This policy, consequently, draws invisible divides within society both in terms of the distribution and accessibility of such products.

6. Conclusion

The frequent use of English in advertisements, notably of Polish brands, indexes a high prestige of the English language in Poland. Its use as well as understanding undoubtedly enhance the status of English-speaking consumers too. The analysed material shows that the two languages, Polish and English, appear side by side in the investigated texts, often marked

off by different fonts and performing certain functions (e.g. expressing slogans) in the genre. However, at times no special multimodal devices are used to separate the two languages, which blend in the same sentence in natural way without any effort to separate the two. As a result, in such cases it is hard to draw clear boundaries between what is native and what is foreign, as both languages appear to be treated as unmarked choices, and the distinction between them is often additionally blurred by an interaction between the text of the advertisement and the information about the product provided on the accompanying image. A prediction may, therefore, be risked that this hybridization of a kind visible in advertising texts may develop into an unspoken norm in Poland. It will then remain to be observed whether the commonality of the phenomenon will, in time, lead to an overall acceptance of the use of English in the analysed genre. This, however, may, paradoxically, contribute to a gradual loss of the prestige that English is enjoying at the moment.

References

Androutsopoulos, J. 2012. English 'on top': Discourse functions of English resources in the German mediascape. *Sociolinguistic Studies 6*(2): 209–238.

Baumgardner, R.J. 2008. The use of English in advertising in Mexican print media. *Journal of Creative Communications* 3(1): 23–48.

Bhatia, T. K., and W. C. Ritchie. 2004. Bilingualism in the global media and advertising. In T.K. Bhatia, and W.C. Ritchie (eds.), *The handbook of bilingualism*, 513–546. Blackwell Publishing.

Blommaert, J. 2010. *The sociolinguistics of globalization*. Cambridge University Press.

Blommaert, J. 2013. *Ethnography, superdiversity and linguistic landscapes. Chronicles of complexity*. Multilingual Matters.

Bulawka, H.M. 2006. *English in Polish advertising*. M.A. dissertation. University of Birmingham. https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/appliedlinguistics/BulawkaDissertation.pdf.

Cheshire, J., and Moser, L.-M. 1994. English as a cultural symbol: The case of advertisements in French-speaking Switzerland. *Journal of Multilingual and Multicultural Development 15*(6): 451–469.

Chłopicki, W., and J. Świątek. 2000. *Angielski w polskiej reklamie*. Wydawnictwo Naukowe PWN.

Collins English Dictionary, https://www.collinsdictionary.com.

Dąbrowska, M. 2020. Multilingual landscape of Kraków. In M. Szczyrbak, and A. Tereszkiewicz (eds.), *Languages in contact and contrast. A festschrift for Professor Elżbieta Mańczak-Wohlfeld on the occasion of her 70th birthday,* 121–132. Jagiellonian University Press.

Eastman, C. M., and Stein, R.F. 1993. Language display: Authenticating claims to social identity. *Journal of Multilingual and Multicultural Development 14*(3): 17–202.

Gardner-Chloros, P. 2009. *Code-switching*. Cambridge University Press.

Garrett, P. 2010. *Attitudes to language*. Cambridge University Press.

Gerritsen, M., Korzilius, H., Van Meurs, F., and I. Gijsbers. 2000. English in Dutch commercials: Not understood and not appreciated. *Journal of Advertising Research 40*(3): 17–31.

Gerritsen, M., Nickerson, C., Crijns, R., van Hooft, A., van Meurs, F., van Nederstigt, U., and M. Starren. 2007. English in product advertisements in Belgium, France, Germany, the Netherlands and Spain. *World Englishes 26*(3): 291–315.

Gottlieb, H. 2005. Anglicisms and translation. In G. Anderman, and M. Rogers (eds.), *In and out of English. For better, for worse,* 161–184. Multilingual Matters.

Grant, R., and J. R. Short (eds.) 2002. *Globalization and the margins*. Palgrave Macmillan.

Griffin, J. 1997. Global English invades Poland: An analysis of the use of English in Polish magazine advertisements. *English Today 13*(2): 34–41.

Griffin, J. 2001. Global English infiltrates Bulgaria. *English Today 17*(4): 54–60.

Grin, F. 1994. The bilingual advertising decision. *Journal of Multilingual and Multicultural Developments 15*(2-3): 269–292.

Haarmann, H. 1989. *Symbolic values of foreign language use: From a Japanese case to a general sociolinguistic perspective*. Walter de Gruyter.

Haugen, E. 1950. The analysis of linguistic borrowing. *Language 26*(2): 211–231.

Hudson, R. 1996. *Sociolinguistics*. 2nd edition. Cambridge University Press.

Kelly-Holmes, H. 1998. The discourse of western marketing professionals in central and eastern Europe. Their role in the creation of a context for marketing and advertising messages. *Discourse and Society 9*(3): 339–362.

Kelly-Holmes, H. 2000. Bier, parfum, kaas: Language fetish in European advertising. *European Journal of Cultural Studies 3*: 67–82.

Kelly-Holmes, H. 2005. *Advertising as multilingual communication*. Palgrave MacMillan.,

Kotler, P., Armstrong, G., Saunders, J., and V. Wong. 2002. *Principles of marketing*. 3rd European edition. Prentice Hall.

Landry, R., and R.Y. Bourhis. 1997. Linguistic landscape and ethnolinguistic vitality. An empirical study. *Journal of Language and Social Psychology 16*(1): 23–49.

Lee, M. 2000. Codeswitching in media texts: Its implications on society and culture in postcolonial Hong Kong. In D. C. S. Li, A. Lin, and W.-K. Tsang (eds.), *Language and education in post-colonial Hong Kong*. Linguistic Society of Hong Kong, 95–127. *Merriam-Webster's collegiate dictionary,* https://www.merriam-webster.com.

Mesthrie, R. 2000. Clearing the ground: Basic issues, concepts and approaches. In R. Mesthrie, J. Swann, A. Deumert, and W.L. Leap (eds.), *Introducing sociolinguistics*, 1–43. Edinburgh University Press.

de Mooij, M. 1994. *Advertising worldwide: Concepts, theories and practice of international, multinational and global advertising*. Prentice Hall.

Muysken, P. 2000. *Bilingual speech: a typology of code-mixing*. Cambridge University Press.

Myers-Scotton, C. 2009. Code-switching. In N. Coupland, and A. Jaworski (eds.), *The new sociolinguistics reader,* 473–489. Palgrave Macmillan.

Piller, I. 2001. Identity construction in multilingual advertising. *Language in Society 30*(4): 153–186.

Piller, I. 2003. Advertising as a site of language contact. *Annual Review of Applied Linguistics 23:* 170–183.

Planken, B., van Meurs, F., and A.Radlińska. 2010. The effects of the use of English in Polish product advertisements: Implications for English for business purposes. *English for Specific Purposes 29*(4): 225–242.

Reklama powinna być prezentowana w takim języku, aby była zrozumiała dla wszystkich odbiorców. *Rzeczpospolita*, 6 November, 2014. https://www.rp.pl/konsumenci/art12178671-reklama-powinna-byc-prezentowana-w-takim-jezyku-aby-byla-zrozumiala-dla-wszystkich-odbiorcow.

Poplack, S. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics* 18: 581–618

Poplack, S. 1988. Contrasting patterns of code-switching in two communities. In M. Heller (ed.), *Codeswitching: Anthropological and sociolinguistic perspectives,* 215–245. Mouton de Gruyter.

Sankoff, G. 2001. Linguistic outcomes of language contact. In J. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The handbook of language variation and change,* 618–668. Blackwell.

Sebba, M. 2012. Multilingualism in written discourse: An approach to the analysis of multilingual texts. *International Journal of Bilingualism 17*(1): 97–118.

Silverstein, M. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication 23*: 193–229.

Spolsky, B. 2009. Prolegomena to a sociolinguistic theory of public signage. In E. Shohamy, and D. Gorter (eds.), *Linguistic landscape. Expanding the scenery,* 25–39. Routledge.

The Act on the Polish language (from October 7, 1999) https://ospcom.files.wordpress.com/2011/11/dr19.pdf.

The Linguistic Landscape: An International Journal (2015). https://www.jbe-platform.com/content/journals/22149961.

Tobolova, B. 2012. *Język angielski w polskiej reklamie telewizyjnej (na podstawie wybranych reklam)*. M.A. dissertation. Masaryk University. https://is.muni.cz/th/fe8tu/Jezyk_angielski_w_polskiej_reklamie_telewizyjnej_praca_dyplomowa.pdf

Trudgill, P. 2003. *A glossary of sociolinguistics*. Edinburgh University Press.

Ustinova, I. 2006. English and emerging advertising in Russia. *World Englishes 25*(2): 267–277.

Ustinova, I. 2008. English and American culture appeal in Russian advertising. *Journal of Creative Communications 3*(1): 77–98.

Ustinova, I., and T. K. Bhatia. 2005. Convergence of English in Russian TV commercials. *World Englishes 24*(4): 495–508.

Vestergaard, T., and K. Schroder. 1985. *The language of advertising*. Blackwell.

Ye, X., and X. Qin. 2004. Forms and functions of Chinese-English mixing in newspaper advertisements of the Guangzhou region. *Journal of South China University of Technology 6*(3): 58–63.

Zhang, H. 2001. An analysis of TV advertising language across cultures. *Studies in the Linguistic Sciences 31*(2): 187–211.

LingBaW *Linguistics Beyond And Within 8 (2022), 55-74*

Spelling out of scope taking arguments in (de-)verbal constructions in Hungarian*

Judit Farkas and Gábor Alberti
University of Pécs, Hungary

Abstract
The paper systematically investigates operators in post-head positions within the three constructions referring to states of affairs in Hungarian, that is, within verbal, deverbal nominal and infinitival phrases. Hungarian is well-known to be a language in which all types of operator can be, and are usually, spelt out in the pre-head zone. However, it has not been discussed in a systematic and comprehensive way earlier whether operators can appear in post-head positions. The paper points out that this is partially possible via a systematic overview of six basic operator types. It also illustrates that while spelling out operators in the pre-head zone results in unambiguous constructions, placing them in post-head positions yields different types of ambiguity. As for the acceptability of scope taking arguments, finite verbal constructions show a black-and-white picture while infinitival and deverbal nominal constructions can be characterized by gray zones in respect of the readiness of arguments to take scope from post-head position. In these "gray zones", a somewhat speaker-dependent variation can be observed, presumably with underlying microvariation. To represent and interpret our findings, we use Grohmann's (2000, 2003) phase-theoretic approach with its pragmasemantics-based three Spell-Out domains per cycle.

Keywords: (de-)verbal phrases, Hungarian, operators, Spell-Out positions

1. Introduction

The paper investigates operators in post-head positions in different verb-based cycles in Hungarian. Hungarian is a language in which all types of operator can be, and are usually, spelt out in the pre-head zone of corresponding cycles, that is, in a Grohmann-style Operator Domain, ΩΔ (Grohmann 2000, 2003, 2009, Farkas and Alberti 2017). As pointed out by É. Kiss (2002: 113–115), the order of operators in the pre-head zone corresponds to scope hierarchy, which guarantees unambiguity in the case of finite verbs (1).[1]

* The paper is supported by the following project: NKFI FK-128518 (The Syntax of Hybrid Categories).

[1] The following abbreviations are used in the glosses:
(i) case suffixes: ACC(USATIVE), DAT(IVE), ILL(ATIVE);
(ii) other suffixes on nouns: POSS (possessedness suffix), POSS.1SG... (possessive agreement suffixes);
(iii) affixes on verbs and infitives: 1SG... (agreement suffix);

(1)

a. [*Minden kollégámat*]$_{\text{Quantifier}}$ [*többször* (*is*)]$_{\text{Quantifier}}$ *meghívtam.*
every colleague.POSS.1SG.ACC several_times also invited.1SG
'It holds for each of my colleagues that I invited him/her several times.'

b. [*Többször* (*is*)]$_{\text{Quantifier}}$ [*minden kollégámat*]$_{\text{Quantifier}}$ *meghívtam.*
several_times also every colleague.POSS.1SG.ACC invited.1SG
'I invited all of my colleagues (together) several times.'

Thus, if the order of the two quantifiers in (1a) is reversed, the meaning changes parallel to the change in scope hierarchy (1b). In (1a), "I could invite my colleagues separately," while in (1b), there had to be more than one occasion when "my colleagues were in my house together."

The same holds for infinitives (2a+b-b')[2] and deverbal nominals (2a+c-c'): operators can appear in the pre-head zone, with word order corresponding to scope hierarchy.[3]

(2)

a. *Na például* […], *az túlzás volt.*
well for_instance that exaggeration was
'Well for instance, as for […] that was an exaggeration.'

b. [*minden kollégámat*]$_{\text{Quantifier}}$ [*többször* (*is*)]$_{\text{Quantifier}}$ *meghív-ni*
every colleague.POSS.1SG.ACC several_times also invite-INF
'inviting each of my colleagues several times'

b'. [*többször* (*is*)]$_{\text{Quantifier}}$ [*minden kollégámat*]$_{\text{Quantifier}}$ *meghív-ni*
several_times also every colleague.POSS.1SG.ACC invite-INF
'inviting all of my colleagues (simultaneously) several times'

c. [*minden kollégámnak*]$_{\text{Quantifier}}$ *a* [*többszöri*]$_{\text{Quantifier}}$ *meghív-ás-a*
every colleague.POSS.1SG.DAT the several_times.ADJ invite-NOM-POSS.3SG
'the invitation of each of my colleagues several times'

c'. [*többször* (*is*)]$_{\text{Quantifier}}$ [*minden kollégámnak*]$_{\text{Quantifier}}$ *a meghív-ás-a*
several_times.ADJ also every colleague.POSS.1SG.DAT the invite-NOM.POSS.3SG
'the invitation of all of my colleagues (simultaneously) several times'

(iv) derivational suffixes: ADJ(EVTIVALIZER), INF(INITIVE), NOM(INALIZER);

Throughout the whole paper, the following six-degree scale of grammaticality judgments, given in Broekhuis, Keizer, and Dikken (2012: viii), is used: *: unacceptable, $^{*?}$: relatively acceptable compared to *; $^{??}$: intermediate or unclear status; $^{?}$: marked: not completely unacceptable or disfavored form; $^{(?)}$: slightly marked, but probably acceptable. We also follow Broekhuis, Keizer, and Dikken (2012: xiv) in using introspective judgements by the group of the two authors (both native speakers of Hungarian) as the criterion of what word orders are part of the language associated with what readings (cf. Featherstone 2007, section 5.4). It is to be noted that, given the peripheral status of certain examples, their evaluation is inevitably speaker-dependent. One of the reviewers, for instance, has given his/her own judgments concerning the data presented in Tables 1–4, which tend to show an at most one-degree difference from ours, without modifying the orderings in Tables 3 and 4 among the six operator types in respect of acceptability.

[2] For the sake of brevity, we often use a shared finite sentential context with several different filler constructions. In formula '(2a+b-b')', for instance, (2a) provides the shared sentential context, in which (2b) and then (2b') should be inserted.

[3] There are two exceptions: *wh*-words can only appear in finite verbal constructions and *also*-quantifiers (containing the particle *is* 'also') cannot appear between D and N heads (for independent reasons), see ex. (766) in Alberti and Farkas (2018: 810).

If the operators appear postverbally, however, the sentence becomes ambiguous. Sentence (3) can be associated with both meanings presented in (1a) and (1b), that is, it is possible in this case that "I invited my colleagues separately or together."

(3) *Meghívtam* [*minden kollégámat*]$_{\text{Quantifier}}$ [*többször* (*is*)]$_{\text{Quantifier}}$.
invited.1SG every colleague.POSS.1SG.ACC several_times also
available meanings: both (1a) and (1b)

In the case of infinitives and deverbal nominals, another kind of ambiguity emerges. These constructions are ambiguous even in the case of a single post-head operator, due to the fact that their scope can be interpreted relative to either the embedded verb (referred to as ISR, 'internal-scope reading') or the finite verb (ESR: 'external-scope reading'). It must be noted that the stress patterns of the two variants are not the same (cf. Varga 2016).

(4)
a. *Hiba volt el-bocsát-ani / el-bocsát-an-od*[4] [*mindkét informatikust*].
mistake was away-allow-INF / away-allow-INF-2SG both IT_specialist.ACC
(?)ISR: 'It was a mistake (for you) to (simultaneously) dismiss [both IT specialists].'
✓ESR: 'In the case of [both IT specialists], it was a mistake (for you) to dismiss each of them.'
b. *Hiba volt az el-bocsát-ás-a* [*mindkét informatikusnak*].
mistake was the away-allow-NOM-POSS.3SG both IT_specialist.DAT
(?)ISR: 'Dismissing (simultaneously) [both IT specialists] was a mistake.'
✓ESR: 'In the case of [both IT specialists], dismissing each of them was a mistake.'

It is also possible to combine the two kinds of ambiguity, resulting in multiple ambiguous constructions (with some of the interpretations undoubtedly being more difficult to retrieve than others), as illustrated in (5). The constructions in question have six different readings due to the fact that both quantifiers can be interpreted internally ("only an ambitious amount of participation in different conferences is contraindicated") as well as externally ("numerous variants of participation in conferences are contraindicated").

(5)
a. *Ellenjavalt el-küld-eni / el-küld-en-ed* [*mindkét kollégát*]
contraindicated away-send-INF / away-send-INF-2SG both colleague.ACC
[*Lublinba is*].
Lublin.ILL also
'It is contraindicated (for you) to send [both colleagues] [also to Lublin].'
b. *Ellenjavalt az el-küld-és-e* [*mindkét kollégának*]
contraindicated the away-send-NOM-POSS.3SG both colleague.DAT
[*Lublinba is*].
Lublin.ILL also
'Sending [both colleagues] [also to Lublin] is contraindicated.'
✓ISR1: [CONTRAINDICATED > BOTH COLLEAGUES > ALSO TO LUBLIN]
??ISR1: [CONTRAINDICATED > ALSO TO LUBLIN > BOTH COLLEAGUES]

[4] Note that in Hungarian infinitives can bear agreement suffixes: the infinitival head agrees with its subject-like argument in number and person in certain cases. Hence, both types will be tested in each construction throughout the whole article.

[??]ESR-ISR1: [BOTH COLLEAGUES > CONTRAINDICATED > ALSO TO LUBLIN]
[?]ESR-ISR2: [ALSO TO LUBLIN > CONTRAINDICATED > BOTH COLLEAGUES]
[??]ESR1: [BOTH COLLEAGUES > ALSO TO LUBLIN > CONTRAINDICATED]
[(?)]ESR2: [ALSO TO LUBLIN > BOTH COLLEAGUES > CONTRAINDICATED]

In the following, the paper concentrates on the mere post-head appearance of different types of operator (without considering constructions containing two or more operators). It has not been systematically investigated in the Hungarian literature whether the different types of operator can appear in a post-head position. In Section 1, it is systematically overviewed in the case of finite verbs, infinitives and deverbal nominals (cf. Alberti and Laczkó 2018) which of the following six basic operator types can appear in a post-head position: *each*-quantifiers, *also*-quantifiers, *only*-focus, negative focus, negative universal quantifier, *wh*-words. Section 2 is devoted to the generalizations of the findings and their interpretation in the Grohmannian (2000, 2003, 2009) phase-theoretical minimalist framework.

2. Operators in post-head positions in Hungarian

This section provides an overview of which operator types can appear in post-verbal, post-nominal and post-infinitival positions in Hungarian; see the (a')-, (b)- and (c)-examples throughout the whole section, respectively. In each series of examples, a finite sentence containing the given operator in question in a pre-verbal position is also provided to present the default position of the operator, see the (a)-examples. If a post-head operator is embedded in a 'for instance'-construction, as in the (b')- and (c')-examples, it tends to become more acceptable. This can be due to the fact that, in this "isolated" situation, the borders of the deverbal construction (i.e., the formula *na például* from left and the resumptive pronoun *az* from right) are clearer, and the scope of the given operator can only be interpreted relative to the embedded verb (Farkas and Alberti 2018: 668–669).

The first operator type to examine is that of *each*-quantifiers. As illustrated in (6a-a'), universal quantifiers can readily appear both in a pre-verbal and in a post-verbal position. The deverbal nominal and infinitival constructions containing a universal quantifier in a post-head position are also highly acceptable (6b-c'), but, in these cases, the sentences are ambiguous due to the fact discussed in the comments concerning (4) above: the scope of the quantifier can be interpreted relative to either the embedded verb or the finite verb.

(6)
a. [*Mindkét informatikust*] *el-bocsátották.*
both IT_specialist.ACC away-allowed.3PL
'[Both IT specialists] were dismissed.'
a'. *El-bocsátották* [*mindkét informatikust*].
away-allowed.3PL both IT_specialist.ACC
'[Both IT specialists] were dismissed.'
b. *Hiba volt az el-bocsát-ás-a* [*mindkét informatikusnak*].
mistake was the away-allow-NOM-POSS.3SG both IT_specialist.DAT
[(?)]ISR: 'Dismissing (simultaneously) [both IT specialists] was a mistake.'
[✓]ESR: 'In the case of [both IT specialists], dismissing each of them was a mistake.'

b'. *Na például az el-bocsát-ás-a* [*mindkét informatikusnak*]
well for_instance the away-allow-NOM-POSS.3SG both IT_specialist.DAT
az hiba volt.
that mistake was
ISR: 'Well for instance, dismissing (simultaneously) [both IT specialists], that was a mistake.'

c. *Hiba volt el-bocsát-ani / el-bocsát-an-od* [*mindkét informatikust*].
mistake was away-allow-INF / away-allow-INF-2SG both IT_specialist.ACC
(?)ISR: It was a mistake (for you) to (simultaneously) dismiss [both IT specialists].'
✓ESR: 'In the case of [both IT specialists], it was a mistake (for you) to dismiss each of them.'

c'. *Na például el-bocsát-ani / (?)el-bocsát-an-od* [*mindkét informatikust*]
well for_instance away-allow-INF / away-allow-INF-2SG both IT_specialist.ACC
az hiba volt.
that mistake was
✓/(?)ISR: 'Well for instance, (for you) to dismiss (simultaneously) [both IT specialists] was a mistake.'

The same holds for quantifiers with *also*: they can appear both in a pre-verbal and in a post-verbal position (7a-a'), and they can appear in the postnominal and in the postinfinitival zone as well; in the latter cases the constructions are ambiguous again in the same way as in the case of universal quantifiers. As illustrated by the comparison between (7b) and (7b'), embedding the *also*-quantifier in a 'for instance'-construction renders the internal-scope reading radically more acceptable.

(7)

a. [*Petit is*] *el-bocsátották.*
Peti.ACC also away-allowed.3PL
'[Peti] was dismissed, [too].'

a'. *El-bocsátották* [*Petit is*].
away-allowed.3PL Peti.ACC also
'[Peti] was dismissed, [too].'

b. (?)*Hiba volt az el-bocsát-ás-a* [*Petinek is*].
mistake was the away-allow-NOM-POSS.3SG Peti.DAT also
??ISR: 'Dismissing [Peti, too], was a mistake.'
(?)ESR: 'In the case of [Peti, too], dismissing him was a mistake.'

b'. (?)*Na például az el-bocsát-ás-a* [*Petinek is*], *az hiba volt.*
well for_instance the away-allow-NOM-POSS.3SG Peti.DAT also that mistake was
(?)ISR: 'Well for instance, dismissing [Peti, too], that was a mistake.'

c. *Hiba volt el-bocsát-ani / el-bocsát-an-od* [*Petit is*].
mistake was away-allow-INF / away-allow-INF-2SG Peti.ACC also
✓ISR: 'It was a mistake (for you) to dismiss [also Peti].'
(?)ESR: 'In the case of [Peti, too], it was a mistake (for you) to dismiss him.'

c'. *Na például el-bocsát-ani / (?)el-bocsát-an-od* [*Petit is*], *az hiba volt.*
well for_instance away-allow-INF / away-allow-INF-2SG Peti.ACC also that mistake was
✓/(?)ISR: 'Well for instance, (for you) to dismiss [also Peti], that was a mistake.'

In the series of examples in (8) focus constructions with the particle *csak* meaning 'only' are investigated. Focus constructions can readily appear preverbally (8a). The presence of focus obligatorily triggers inverse word order in Hungarian, that is, the finite verb stem must precede the preverb. In contrast to the two quantifier types tested so far, focus constructions

cannot appear postverbally, neither with a [verb stem + preverb], nor with a [preverb + verb stem] word order, as shown in (8a').

(8)

a. [*Csak Petit*] [*bocsátották el*] / **el-bocsátották.*
only Peti.ACC allowed.3PL away / away-allowed.3PL
'[Only Peti] was dismissed.'

a'. **El-bocsátották* / *[*Bocsátották el*] [*csak Petit*].
away-allowed.3PL / allowed.3PL away only Peti.ACC
Intended meaning: '[Only Peti] was dismissed.'

b. ??*Hiba volt az el-bocsát-ás-a* [*csak Petinek*].
mistake was the away-allow-NOM-POSS.3SG only Peti.DAT
??ISR: 'Dismissing [only Peti] was a mistake.'
*Intended ESR: 'It is [only Peti] whose dismissal was a mistake.'

b'. ?*Na például az el-bocsát-ás-a* [*csak Petinek*], *az hiba volt.*
well for_instance the away-allow-NOM-POSS.3SG only Peti.DAT that mistake was
?ISR: 'Well for instance, dismissing [only Peti], that was a mistake.'

c. ??*Hiba volt el-bocsát-ani / el-bocsát-an-od* [*csak Petit*].
mistake was away-allow-INF / away-allow-INF-2SG only Peti.ACC
??ISR: 'It was a mistake (for you) to dismiss [only Peti].'
*Intended ESR: 'It is in the case of [only Peti] that it was a mistake (for you) to dismiss him.'

c'. *Na például* ?*el-bocsát-ani / ??el-bocsát-an-od* [*csak Petit*], *az hiba volt*
well for_instance away-allow-INF / away-allow-INF-2SG only Peti.ACC that mistake was
?/??ISR: 'Well for instance, to dismiss [only Peti], that was a mistake.'

Deverbal nominal constructions are somewhat more permissive with a postnominal *only*-focus (8b). Note, however, that external-scope reading cannot be associated with constructions like this, that is, the scope of the focus must be interpreted relative to the embedded verb in cases like this. If a postnominal *only*-focus is embedded in a 'for instance'-construction, as in (8b'), it becomes more acceptable. This can be due to the fact that, in this case, the borders of the deverbal nominal construction are clearer, and the scope of the focus can only be interpreted relative to the embedded verb. As illustrated in (8c-c'), the same holds for infinitival constructions as well.

Negative focus constructions are similar to *only*-focus constructions: they can only appear preverbally (compare (9a) and (9a')) and they also trigger inverse word order in Hungarian. What makes negative focus constructions different from *only*-focus constructions is that they are also unacceptable in postnominal and postinfinitival constructions, as shown in (9b-b',c-c'). They are somewhat more acceptable embedded in a 'for instance'-construction, but those constructions do not reach a convincingly acceptable level, either.

(9)

a. [*Nem Petit*] [*bocsátották el*] / **el-bocsátották.*
not Peti.ACC allowed.3PL away / away-allowed.3PL
'[Not Peti] was dismissed.'

a'. **El-bocsátották* / *[*Bocsátották el*] [*nem Petit*].
away-allowed.3PL / allowed.3PL away not Peti.ACC
Intended meaning: 'It was [not Peti] who was dismissed.'

b. *?*Hiba volt az el-bocsát-ás-a* [*nem Petinek*].
mistake was the away-allow-NOM-POSS.3SG not Peti.DAT
*?*Intended ISR: 'It was a mistake to dismiss [not Peti but someone else].'
*Intended ESR: 'It is [not Peti] whose dismissal was a mistake.'

b'. *?*Na például az el-bocsát-ás-a* [*nem Petinek*], *az hiba volt.*
well for_instance the away-allow-NOM-POSS.3SG not Peti.DAT that mistake was
*?*ISR: 'Well for instance, dismissing [not Peti but someone else], that was a mistake.'

c. *?*Hiba volt el-bocsát-ani / el-bocsát-an-od* [*nem Petit*].
mistake was away-allow-INF / away-allow-INF-2SG not Peti.ACC
*?*Intended ISR: 'It was a mistake (for you) to dismiss [anyone else than Peti].'
*Intended ESR: 'It is in the case of [anyone else than Peti] that it was a mistake (for you) to dismiss him or her.'

c'. *Na például *?*el-bocsát-ani / *el-bocsát-an-od* [*nem Petit*],
well for_instance away-allow-INF / away-allow-INF-2SG not Peti.ACC
az hiba volt.
that mistake was
*?/**Intended ISR: 'Well for instance, (for you) to dismiss [anyone else than Peti], that was a mistake.'

A *wh*-word can only appear in a preverbal position in Hungarian triggering inverse word order (10a), but it cannot appear in any post-head positions (10a',b-c').[5] The only exceptions are the cases in which a *wh*-word also appears in a preverbal position (10a").

(10)

a. [*Kit*] [*bocsátottak el*] / **el-bocsátottak*?
who.ACC allowed.3PL away / away-allowed.3PL
'[Who] was dismissed?'

a'. **El-bocsátottak* / *[*Bocsátottak el*] [*kit*]?
away-allowed.3PL / allowed.3PL away who.ACC
Intended meaning: '[Who] was dismissed?'

a". [*Kit*] *bocsátottak el* [*honnan*]?
who.ACC allowed.3PL away from_where
'[Who] was dismissed [from where]?'

b. **Hiba volt az el-bocsát-ás-a* [*kinek*]?
mistake was the away-allow-NOM-POSS.3SG who.DAT
*ISR: -
*ESR: -

b'. **Na például az el-bocsát-ás-a* [*kinek*], *az hiba volt.*
well for_instance the away-allow-NOM-POSS.3SG who.DAT that mistake was
*ISR: -

c. **Hiba volt el-bocsát-ani / el-bocsát-an-od* [*kit*]?
mistake was away-allow-INF / away-allow-INF-2SG who.ACC
*ISR: -
*Intended ESR: 'Whom does it hold for that it was a mistake (for you) to dismiss?'

c'. *Na például *?*el-bocsát-ani / *el-bocsát-an-od* [*kit*], *az hiba volt?*
well for_instance away-allow-INF / away-allow-INF-2SG who.ACC that mistake was
*ISR: -

[5] In echo-questions, interrogatives can appear postverbally with a special intonation. Due to the fact that such questions have a semantic/pragmatic status different from that of the ones considered in (10), we postpone their investigation to future research.

The last operator type to test is the group of negative universal quantifiers. It is exemplified in (11) that negative universal quantifiers can readily appear in post-head positions. As can be seen in (11a'), if a quantifier like this appears postverbally, the negative particle *nem* 'not' must appear before the finite verb stem (triggering an inverse word order), while another negative particle can also appear in the construction. The appearance of *sem* 'also not', however, is optional in this case (NB: the last consonant 'm' in the particle *sem* is always optional; its appearance depends on the register used by the speaker). If the negative universal quantifier appears preverbally (11a), both *sem* and *nem* can be used, but not simultaneously.

(11)

a. [*Senkit*] *se*(*m*) / *nem* *bocsátottak* *el.*
no-one.ACC not / not allowed.3PL away
'[No-one] was dismissed.'

a'. *Nem* *bocsátottak* *el* [*senkit* (*se*(*m*))].
not allowed.3PL away no-one.ACC not
'[No-one] was dismissed.'

b. (?)*Nem* *volt* *hiba* *az* *el-bocsát-ás-a* [*senkinek* (*se*(*m*))].
not was mistake the away-allow-NOM-POSS.3SG no-one.DAT not
(?)ESR: 'It holds for [everyone] that dismissing him/her was not a mistake.'

b'. *Na* *például* *az* *el* **(?)nem*) *bocsát-ás-a* [*senkinek* (*se*(*m*))],
well for_instance the away not allow-NOM-POSS.3SG no-one.DAT not
az *hiba* *volt.*
that mistake was
(?)ISR: 'Well for instance, dismissing [no-one], that was a mistake.'

Let us consider the corresponding deverbal nominal constructions in (11b-b'). They are acceptable but note that (11b) is not ambiguous: it can only be associated with external-scope reading. It is illustrated in (11b'), which is restricted to the internal-scope reading due to the test construction, that this reading is available only if the negative particle *nem* appears before the deverbal nominal (NB: the word-order alternatives among the preverb, the deverbal nominal and the negative particle, the order given here is the most acceptable one).

As shown in (12), there are many word-order variants in the case of infinitives. This is due to the fact that the negative particle *nem*, the preverb *el* and the infinitival head *bocsátani* can appear in three different orders: [Neg(ative)P(ar)t(i)c(le)+Inf+PreV], [NegPtc+PreV+Inf] or [PreV+NegPtc+Inf].

(12)

a. *Hiba* *volt* [...].
mistake was
'It was a mistake [...].'

b. *nem* *bocsát-ani* *el* [*senkit* (*se*(*m*))]
not send-INF away no-one.ACC not

b'. *nem* *el-bocsát-ani* [*senkit* (*se*(*m*))]
not away-send-INF no-one.ACC not

b". *el* *nem* *bocsát-ani* [*senkit* (*se*(*m*))]
away not send-INF no-one.ACC not
'to dismiss [no-one]'

c. *nem bocsát-an-od el* [*senkit* (*se*(*m*))]
not send-INF-2SG away no-one.ACC not
c'. *nem el-bocsát-an-od* [*senkit* (*se*(*m*))]
not away-send-INF-2SG no-one.ACC not
c". *el nem bocsát-an-od* [*senkit* (*se*(*m*))]
away not send-INF-2SG no-one.ACC not
'for you to dismiss [no-one]'

In Table 1 below, the x/y pairs of grammaticality judgments belong to the variants without and with *sem* 'not' in (12), respectively. The table clearly shows that the grammaticality judgments associated with the constructions are not the same even in the case of the two authors (both native speakers of Hungarian). For instance, some Hungarians (like speaker1) tolerate the [NegPtc+Inf+PreV] order well also in inflected infinitival constructions, while others (like speaker2) do not (see also Farkas 2020). However, in uninflected infinitival expressions, both speakers tolerate the 'preverb last' order well. The underlying microvariations behind this whole phenomenon are worth future research. Our analysis can be regarded as a point of departure which can serve as the "null hypothesis" for would-be statistical tests whose aims should be decided on the basis of theory-internal purposes. Note that the constructions in question can only be associated with internal-scope reading (that is, the scope of the negative universal quantifier can only be interpreted relative to the embedded infinitive).

Table 1: *Grammaticality judgments from two speakers pertaining to the test construction given in (12)*

(12)	**b.**	**b'.**	**b".**	**c.**	**c'.**	**c".**
SP1	??/(?)	?/*?	?/*?	?/??	?/*?	?/*?
SP2	?/?	?/?	*/*	*?/*?	?/?	*?/*?

There is another series of examples in (13) concerning the same negative universal quantifiers in the postinfinitival complement zone. In this case, however, the negative particle *nem* precedes the finite verb stem, triggering the external scope reading (NB: the two authors' grammaticality judgments are now similar).

(13)
a. *Nem volt hiba* [...].
not was mistake
'It was not a mistake [...].'
b. *el-bocsát-ani* [*senkit* ?(?)*se*(*m*))]
away-send-INF no-one.ACC not
'to dismiss [anyone]'
c. *el-bocsát-an-od* [*senkit* ??(?)*se*(*m*))]
away-send-INF-2SG no-one.ACC not
'to dismiss [anyone]'

In (14), the same six constructions presented in (12) are tested, but now embedded in a 'for instance'-construction. That is, only the internal scope reading was available in these cases.

(14)

a. *Na például* [...], *az hiba volt.*
well for_instance that mistake was
ISR: 'Well for instance, [...], it was a mistake.'

b. *nem bocsát-ani el* [*senkit* (*se*(*m*))]
not send-INF away no-one.ACC not

b'. *nem el-bocsát-ani* [*senkit* (*se*(*m*))]
not away-send-INF no-one.ACC not

b". *el nem bocsát-ani* [*senkit* (*se*(*m*))]
away not send-INF no-one.ACC not
'to dismiss [no-one]'

c. *nem bocsát-an-od el* [*senkit* (se(m))]
not send-INF-2SG away no-one.ACC not

c'. *nem el-bocsát-an-od* [*senkit* (*se*(*m*))]
not away-send-INF-2SG no-one.ACC not

c". *el nem bocsát-an-od* [*senkit* (*se*(*m*))]
away not send-INF-2SG no-one.ACC not
'for you to dismiss [no-one]'

As can be seen in Table 2, the grammaticality judgments of the two authors are different again. Very similar tendencies can be observed as in the case of Table 1. First, it is only the construction of the uninflected infinitival form in which both speakers tolerate the [NegPtc+Inf+PreV] order well. Second, in inflected infinitival constructions, this 'preverb-last' order is unacceptable to speaker2 while acceptable to speaker1.

Table 2: *Grammaticality judgments from two speakers pertaining to the test construction given in (14)*

(14)	**b.**	**b'.**	**b".**	**c.**	**c'.**	**c".**
SP1	?/ ✓	??/*?	(?)/?	?/??	??/*?	(?)/??
SP2	?/ ✓	?/??	??/*?	*/*	?/??	*/*

In the case of negative universals, instead of two scope hierarchies (the internal and the external one), altogether six scope hierarchies are to be considered. This is due to the fact that negative universals consist of two logical elements: negation and universal quantification, and hence three logical components (mistake in the finite predicative construction is the third one) should be arranged in different orders.

Let us consider the best variant in (12a-b), repeated here as (15):

(15) *Hiba volt nem bocsát-ani el* [*senkit* (*se*(*m*))].
mistake was not send-INF away no-one.ACC not
'It was a mistake to dismiss [no-one].'

The series of examples in (16) present the potential six scope hierarchies. The best Hungarian variants with the logical forms are also provided for all scope orders. As demonstrated, the word-order variant presented in (15) can only be associated with the meaning in (16a). (16a') has the same meaning expressed via an unambiguous construction. Note that the negative universal in (16a') is placed in the preinfinitival zone and, in harmony with its pre-head position, it can only be associated with the intended meaning.

(16) ● Potential scope hierarchies in the case of negative universal quantifiers in InfPs

a. mistake > ∀ > ¬
'It was a mistake that in the case of each person it had been decided that they would not been dismissed.'

a'. *Hiba volt senkit sem el-bocsát-ani.*
mistake was no-one.ACC not away-send-INF

b. mistake > ¬ > ∀
'It was a mistake that it had not be the case that each person would be dismissed.'

b'. *Hiba volt nem mindenkit el-bocsát-ani.*
mistake was not everyone.ACC away-send-INF

c. ∀ > mistake > ¬
'In the case of each person it was a mistake that they had not been dismissed.'

c'. *Mindenkit hiba volt nem el-bocsát-ani.*
everyone.ACC mistake was not away-send-INF

d. ∀ > ¬ > mistake
'In the case of each person it was not a mistake that he had been dismissed.'

d'. *Senkit sem volt hiba el-bocsát-ani.*
no-one.ACC not was mistake away-send-INF

e. ¬ > ∀ > mistake
'It is not the case that in the case of each person it was a mistake that they had been dismissed.'

e'. *Nem mindenkit volt hiba el-bocsát-ani.*
not everyone.ACC was mistake away-send-INF

f. ¬ > mistake > ∀
'It is not the case that it was a mistake that each person had been dismissed.'

f'. *Nem volt hiba mindenkit el-bocsát-ani.*
not was mistake everyone.ACC away-send-INF

The puzzle is that we have six potential variants with different scope hierarchies but the meaning of the construction in (15) with a negative universal in the postinfinitival complement zone can only be associated with the internal meaning, presented in (16a-a').
We offer the following solution to the puzzle.

First of all, the position of the negative particle *nem* 'not' disambiguates between internal/external readings in the following way. If it is adjoined to the finite verbal construction *hiba volt* 'was a mistake', it is only the external scope reading that is available, as in (16d-d'); but if it is adjoined to the embedded (infinitival) verb *elbocsátani* 'dismiss', only the internal scope reading is available, as in (15) and (16a-a').

Second, *senki* 'no-one' determines the scope order between negation and universal quantification: universal quantification always has an immediate scope over negation (17a-a'). In this way, only two of the six variants remain: the first one when the finite verb has a scope over these two, resulting in the internal scope reading (17b), and the second one when universal quantification has a scope over negation, which has a scope over the finite verb, resulting in the external scope reading (17b'). And (again) the position of the negative particle tells us which is the case.

(17)

a. … ∀ > ¬ …

a'. *[… ∀ > X > ¬ …]

b. mistake > ∀ > ¬ → internal scope

b'. ∀ > ¬ > mistake → external scope

c. *nem … s-en-ki … s-e(m)* Szabolcsi (2018: 240)
not also-not-who also-not

There are further factors to mention. The first one is that the optimal word order variant containing a negative universal quantifier in the postinfinitival zone is realized by a "triple negation", schematized in (17c). This variant contains the negative particle *nem* 'not', the negative universal quantifier *senki* 'no-one', and another negative particle *sem* 'also not'. The variant with explicit *sem* is better than the variant without it; compare the grammaticality judgments associated with (12b) in Table 1.

The second factor to mention is that the negative particle *nem* 'not' can appear in the infinitival construction in three ways: [NegPtc+Inf+PreV], [NegPtc+PreV+Inf] or [PreV+NegPtc+Inf]. It is the [NegPtc+Inf+PreV] order (with the same build-up as the finite verbal constructions) that seems to be the best solution (at least in constructions containing non-agreeing infinitives).

To sum up, Table 3 and 4 below present the grammaticality judgments associated with the constructions containing the six operator types tested in post-head zones.

First let us consider the judgments belonging to internal scope (Table 3). In the case of finite verbs, the picture is black and white: the three quantifier types can readily appear postverbally, while it is impossible for the three focus types to appear here. The picture is less black and white in the case of infinitives and deverbal nominals: quantifiers are not always fully acceptable in the post-head zones but constructions containing different types of focus are somewhat more acceptable, except for *wh*-words, which can never have internal scope in the post-head zone.

Table 3: *Readiness of arguments of finite verbs, infinitives and deverbal nominals to take internal scope in the post-head zone*[6]

	Argument type	**mind**	**is**	**se-**	**csak**	**nem**	**wh**
Finite verb	subject	✓	✓	✓	*	*	*
Finite verb	non-subject	✓	✓	✓	*	*	*
Infinitive	non-subject-like	✓/✓	✓/✓	✓/?	?/??	*?/*?	*/*
InfinitiveAgr	non-subject-like	(?)/✓	(?)/✓	?/?	??/??	*/*?	*/*
InfinitiveAgr	subject-like	(?)/(?)	(?)/(?)	(?)/?	??/??	*?/*?	*/*
ás-N	possessor: Theme	✓ / (?)	(?) / ??	? / ??	? / ??	?? / *?	¬* / *
ás-N	non-possessor	✓ / (?)	(?) / ??	? / ??	? / ??	?? / *?	* / *

In the case of external scope, presented in Table 4, the picture is black and white again in the case of finite verbs: quantifiers are fully acceptable in this zone, while the different types of focus can never appear postverbally.

Infinitives and deverbal nominals are similar to finite verbs in that the three types of focus expression cannot appear in the post-head zone, but they also differ from the finite verbs: the

[6] Due to space limitations, three rows of Tables 3 and 4, marked with 'argument type' labels *in italics*, contain data with no exemplification in the paper. The relevant examples are available in the corresponding volumes of *Comprehensive Grammar Resources: Hungarian* (Alberti and Laczkó 2018, Alberti to appear).

three types of quantifier are somewhat less acceptable in post-head zones of deverbal constructions than they are in the case of finite verbs.

Table 4: *Readiness of arguments of finite verbs, infinitives and deverbal nominals to take external scope in the post-head zone*

	Argument type	**mind**	**is**	**se-**	**csak**	**nem**	**wh**
Finite verb	subject	✓	✓	✓	*	*	*
Finite verb	non-subject	✓	✓	✓	*	*	*
Infinitive	non-subject-like	(?)	(?)	(?)	*	*	*
InfinitiveAgr	non-subject-like	(?)	(?)	(?)	*	*	*
InfinitiveAgr	subject-like	?	?	??	*	*	*
ás-N	possessor: Theme	✓	(?)	(?)	*	*	*
ás-N	non-possessor	?	??	??	*	*	*

3. Generalizations of the findings and their interpretation in a phase-theoretic model

Following Farkas and Alberti's (2017) paper on a special Hungarian deverbal nominal construction, we use Grohmann's phase-theoretic approach with its pragmasemantics-based three Spell-Out domains per cycle (Grohmann 2000, 2003, 2009), presented in Figure 1 ("fine-tuned" by Sigurðsson's (2009) context-sensitive Syntax–Phonology interface). The importance of the "pragmasemantic basis" lies in the fact that, for each argument of a head and for each Spell-Out domain of the cycle of the given head, it can be investigated what semantic or pragmatic aspect of the given argument is exhibited in the given domain encoded by the domain-internal position of the copy of the argument (Alberti and Farkas 2021, Farkas and Szabó and Alberti 2021).

In the thematic domain (ΘΔ) of a verbal cycle, the positions in the syntactic hierarchy encode thematic roles in a traditional approach, or event-structural characters. Domain ΦΔ, at first glance, seems to be responsible for case and agreement morphology, and not for pragmasemantic factors, but there are crucial factors in the backgorund, such as linkedness to discourse participants (persons, objects, temporal and local entities) (Doner 2018). In ΩΔ, operator functions assigned to arguments are expressed. The word order of a particular sentence is accounted for in Grohmann's model (similar to other minimalist models) via deciding which argument or adjunct is spelt out in which domain and which verb (or other cycle center) is spelt out in which functional head in the course of its typically long route from head to head upwards in its cycle.

The schematic syntactic tree in Figure 2 of the InfP of sentence (18) illustrates the possibility of spelling out arguments in ΩΔ, due to their operator character: the object and the dative case-marked argument are exhibited as the topic (Top) and the identificational (narrow) focus (Foc) of the given sentence, respectively.

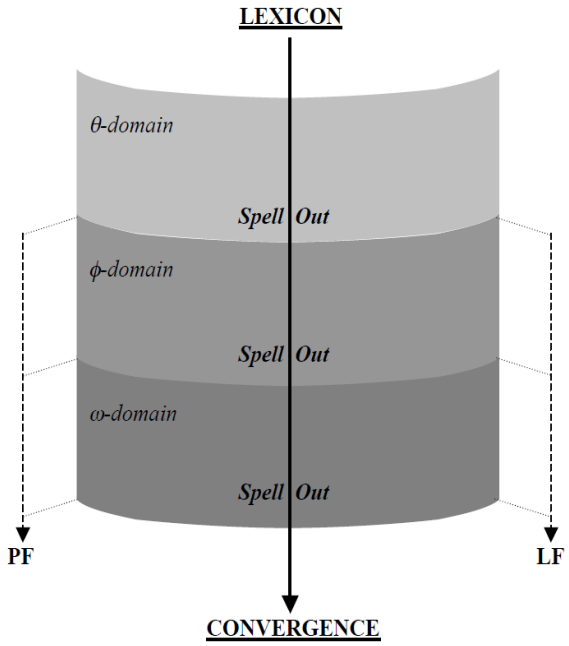


Figure 1: *Spell-out phases in Grohmann's (2000: 291) model*

(18) *Na például Bogáncsot*$_{\text{Topic}}$ *éppen Marcsinak*$_{\text{Focus}}$ *odaadni, az hiba volt.*
well for_instance Bogáncs.ACC just Marcsi.DAT towards.give.INF that mistake was
ISR: 'Well for instance, in the case of Bogáncs, to give him exactly to Marcsi, that was a mistake.'

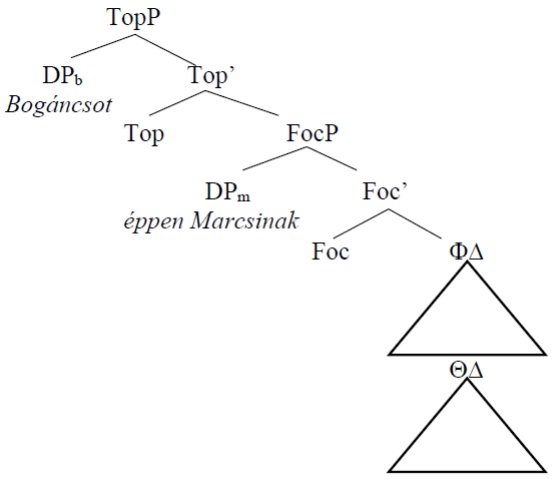


Figure 2: *Schematic illustration of the operator domain of an infinitival phrase*

The six sentence variants presented in (19) below with their slightly different grammaticality judgments are devoted to illustrate instances of spelling out arguments in the lower two domains. Variants (19a) and (19a') show that it is preferred for a Theme to precede the Beneficiary. A phonetically heavy Theme, however, preferably follows the Beneficiary (19b-b'), a phenomenon attributed to Behaghel's Law by É. Kiss (2009). The (c)-examples corroborate these observations: the less acceptable variant is (19c') in which both the preferred thematic order and the Behaghel-based order are violated.

(19) ● Argument ordering on the basis of thematic character or phonetic weight

a. *Hiba lenne odaadni* [*Bogáncsot*]$_{\text{Theme}}$ [*Marcsinak*]$_{\text{Beneficiary}}$.
mistake would_be towards.give.INF Bogáncs.ACC Marcsi.DAT
'It would be a mistake to give Bogáncs to Marcsi.'

a'. (?)*Hiba lenne odaadni* [*Marcsinak*]$_{\text{Beneficiary}}$ [*Bogáncsot*]$_{\text{Theme}}$.
mistake would_be towards.give.INF Marcsi.DAT Bogáncs.ACC
'It would be a mistake to give Marcsi Bogáncs.'

b. (?)*Hiba lenne odaadni* [*azt a kis foltos kutyust*] [*Marcsinak*].
mistake would_be towards.give.INF that.ACC the small spotty puppy.ACC Marcsi.DAT
'It would be a mistake to give that small spotty puppy to Marcsi.'

b'. *Hiba lenne odaadni* [*Marcsinak*]
mistake would_be towards.give.INF Marcsi.DAT
[*azt a kis foltos kutyust*]
that.ACC the small spotty puppy.ACC
'It would be a mistake to give Marcsi that small spotty puppy.'

c. *Hiba lenne odaadni* [*Bogáncsot*]
mistake would_be towards.give.INF Bogáncs.ACC
[*annak a megbízhatatlan elefántcsontparti lánynak*].
that.DAT the unreliable Ivorian girl.DAT
'It would be a mistake to give Bogáncs to that unreliable Ivorian girl.'

c'. ?*Hiba lenne odaadni*
mistake would_be towards.give.INF
[*annak a megbízhatatlan elefántcsontparti lánynak*] [*Bogáncsot*].
that.DAT the unreliable Ivorian girl.DAT Bogáncs.ACC
'It would be a mistake to give that unreliable Ivorian girl Bogáncs.'

The three primeless and the three primed word-order variants can be analyzed as having syntactic structures with arguments spelt out in ΘΔ and in ΦΔ, respectively. The detailed syntactic analysis presented in Figure 3 below illustrates the case of Spell Out in ΦΔ. In this domain, arguments and adjuncts are assumed to be ordered according to Behaghel's Law only to account for the fact that in Hungarian post-head phrases are preferably ordered this way.[7] One might ask what is the aforementioned "pragmasemantic background" behind Behaghel's Law. A plausible answer can be based on the procedure of *anchoring* referents and can be exemplified as follows: In the case of the perfect sentence (19b'), during reading the sentence from left to right, all but one participants could be anchored as soon as at the 11th syllable. In the case of the less acceptable word-order variant given in (19b), however, all but one participant could be anchored only at the 15th syllable.

[7] In the complement of the Asp(ectual) head, the order of the functional heads (Cn 'central' for „distinguished" arguments and Nc 'non-central' for other arguments and adjuncts) whose specifiers host the arguments and adjuncts belonging to the given cycle is taken to be free; Behaghel's Law, hence, works as a filter on the set of potential orderings. This means that the statement according to which the three sentences with [Theme, Beneficiary] order and the three with the reversed order can be analyzed as having syntactic structures with arguments spelt out in ΘΔ and in ΦΔ, respectively, is somewhat simplified. *Any* argument order can be "arranged" and spelt out in ΦΔ; in ΘΔ, however, only a single "thematic order" can be spelt out.

Figure 3: *Detailed Grohmannian syntactic representation of the infinitival phrase odaadnia Marcsinak [azt ... kutyust] 'for him/her to give Marcsi [that ... puppy]', cf. (17b')*

Due to space limitations, we cannot elaborate on further details of the syntactic tree that represents the given infinitival phrase. What is relevant here is that the verbal root *ad* 'give' is climbing upwards from head to head, in the course of which it reaches the N head with the derivative morpheme *-ni* 'to' therein. Their merger produces the infinitival form *adni* 'to give'. Regarding the N head as the center of the InfP, we can say that its complement is the embedded verb's cycle with two domains (ΘΔ, ΦΔ; now there is no ΩΔ) and its immediate projection NP is accommodated in a nominal cycle with one domain (ΦΔ) (Fu *et al.* 2001). As all word-order variants listed in (19) are such that the infinitival head precedes the two arguments, they are spelt out in one of the two verbal domains. In the case of sentence (18), however, the given word order can be accounted for by analyzing the two arguments as spelt out in ΩΔ of the nominal cycle.

If an operator *could* have been spelt out in ΩΔ of the nominal cycle but *is* actually spelt out after the infinitival/deverbal nominal head, as in (20a-b) and in several examples in Section 1, the corresponding syntactic structure should contain a covert movement from ΦΔ to ΩΔ, as illustrated in Figure 4.

(20)

a. *Na például odaadnia Marcsinak mindkét kutyust*$_{\text{Quantifier}}$, *az hiba volt.*
well for_instance towards.give.INF.3SG Marcsi.DAT both puppy.ACC that mistake was
ISR: 'Well for instance, for him/her to give Marcsi both puppies, that was a mistake.'

b. (?)*Hiba lenne odaadni* [*azt a kis foltos kutyust*] [*Marcsinak*].
mistake would_be towards.give.INF that.ACC the small spotty puppy.ACC Marcsi.DAT
'It would be a mistake to give that small spotty puppy to Marcsi.'

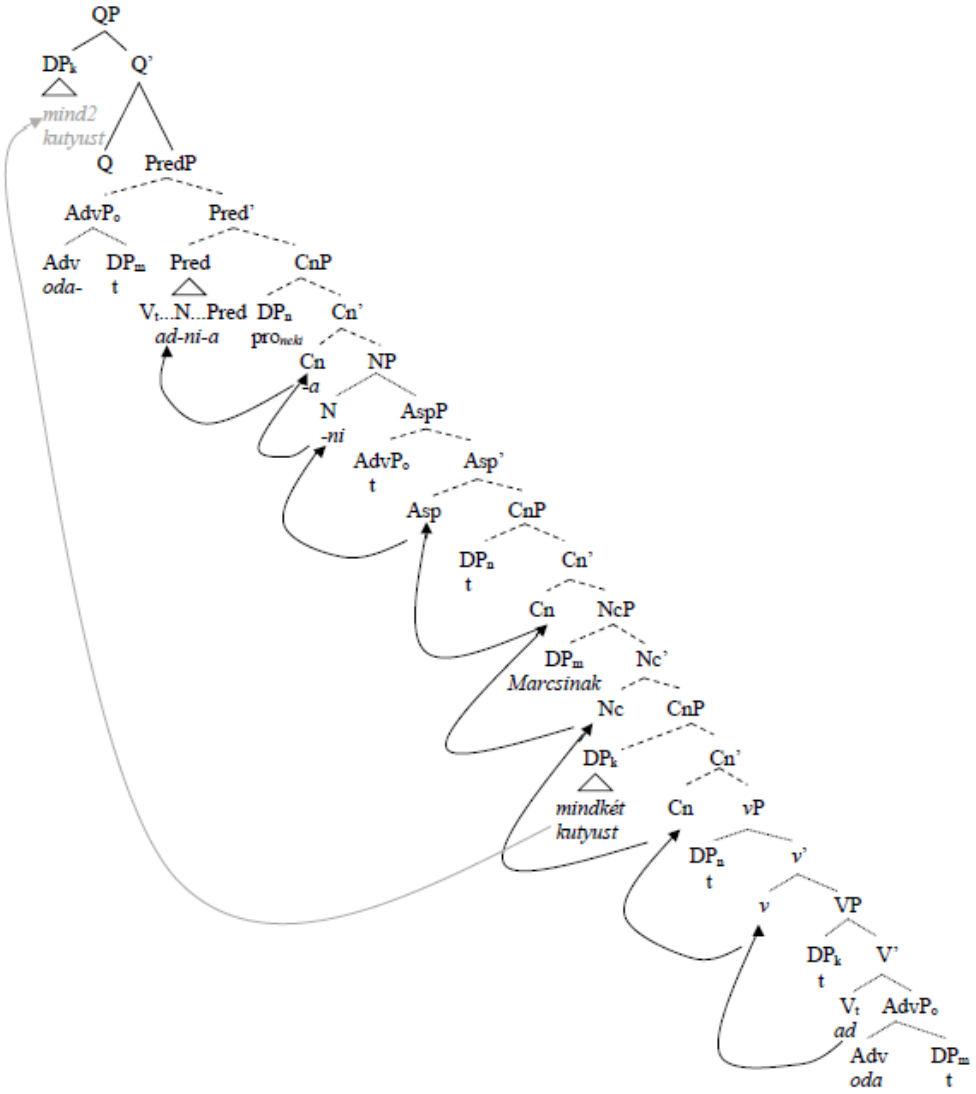


Figure 4: *Post-head appearance of an operator, accounted for by covert movement*

In Figure 4, a quantifier expression is spelt out in ΦΔ, where arguments are ordered according to Behaghel's Law (independent of their +/– scope taking character), instead of being spelt out in ΩΔ. It thus does not make its scope explicit. Its scope is still decided in ΩΔ; however, not overtly but after the point of spelling out.

In (20a), the scope in question is necessarily an internal one. If the same expression were interpreted externally, which is an option in (20b), its syntactic positions in the tree (and their visibility) should be the same, according to Farkas, Szabó and Alberti (2021), see Figures 1 and 3. In that paper (as well as in the 2017 paper of the same three authors), the external interpretation of the given scope is attributed to a kind of feature percolation, proposed in the

Hungarian literature by Horvath (1997: 547–557) and Kenesei (1998: 223–225) in the case of other phenomena. Thus the syntactic difference between the two interpretations simply lies in the position of the operator feature. If the feature remains with the highest copy of DP_k, the interpretation is internal. If, however, it leaves DP_k and percolates up to the node that represents the whole infinitival expression, the interpretation is external. In the given case, both instances of donation are claimed to be a mistake (where one donating event is the donation of one of the puppies, and the other donating event is the donation of the other puppy; NB: on the internal reading, there is a single donating event, in the course of which two puppies are given to someone).

In this light, the observations collected in Tables 3 and 4 in Section 1 can be understood as follows: scope taking operators of different types in Hungarian differ in tolerating being spelt out in ΦΔ and moved up to ΩΔ only covertly, which seems to be the basic fate of operators in the *configurational* English language (É. Kiss 1987). Operator types also differ in the toleration of the kind of feature percolation discussed above (claimed to produce external readings).

4. Summary

What character of an argument can be exhibited via the domain where it is spelt out? – it is in this way that one of the central questions of the paper can be formulated in our Grohmannian approach. Being spelt out in *any* of the three domains can reveal some valuable information on a constituent. However,

- it differs from language to language,
- and even from cycle to cycle within a language,
- and from operator type to operator type in the case of scope taking expressions,

whether the given expression can be spelt out in one or two or all three Grohmannian domains. Or even more precisely, it differs from language to language, cycle to cycle, and operator to operator in what domain the spelling out of the given expression is tolerated to what extent.

As a point of departure, it has been declared as a basic property of Hungarian that all six operator types considered in the paper can be spelt out in ΩΔ of their cycles. The advantage of spelling out scope taking arguments in ΩΔ is that hearers obtain disambiguated scope relations, relative to both co-arguments and the finite verb (if the cycle considered does not happen to be that of the finite verb).

Three operators, the three quantifier types and only they, can *basically* be spelt out in a post-head position, too, that is, spelt out in ΦΔ or ΘΔ (appearing in ΩΔ only covertly). 'Basically' is inserted in this generalization for the following reason. Of the three constructions referring to states of affairs, ones with finite V heads show the generalization as a black-and-white picture. Infinitival and deverbal nominal constructions, as shown in Tables 3 and 4, can be characterized by "gray zones" in respect of the readiness of arguments to take scope covertly. If one makes a comparison between the corresponding rubrics of the two tables, they

can formulate the hypothesis that the mechanism of feature percolation producing external reading is further "cost" somewhat decreasing the acceptability of the syntactic structure. As illustrated in Tables 1 and 2, in these gray zones, there is also a somewhat speaker-dependent variation, presumably with underlying microvariation. We have also pointed out that word order typically does not disambiguate between internal and external readings. The case of negative universal quantifiers is an exception, essentially due to what can be regarded in Hungarian as not only double but triple negation.

Post-head arguments (independent of their +/– scope taking character) are preferably ordered according to Behaghel's Law, which can be attributed to their spelling out in ΦΔ.

References

Alberti, G. (ed.) to appear. *Syntax of Hungarian: Infinitival and Other Non-Finite Constructions.* Amsterdam, The Neatherlands: Amsterdam University Press.

Alberti, G., and J. Farkas. 2018. Modification. In G. Alberti, and T. Laczkó (eds.), *Syntax of Hungarian: Nouns and noun phrases,* 775–895. Amsterdam, The Neatherlands: Amsterdam University Press.

Alberti, G., and J. Farkas. 2021. Igei és hibrid szerkezetek Grohmann-ciklusai és tartományai a magyarban [Grohmannian cycles and domains of verbal and hybrid constructions in Hungarian]. *Általános Nyelvészeti Tanulmányok* [Studies in General Linguistics] XXXIII: 13–42.

Alberti, G., and T. Laczkó (eds.) 2018. *Syntax of Hungarian – Nouns and Noun Phrases*. Amsterdam, The Neatherlands: Amsterdam University Press.

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

Doner, J. 2018. The Anchoring Domain and the EPP. Talk at *GLOW*41. http://seas3.elte.hu/tmp/glowabstracts/glow41-Doner.pdf

É. Kiss K. 1987. Configurationality in Hungarian. Reidel, Dordrecht & Akadémiai Kiadó, Budapest.

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, Great Britain and New York, USA: Oxford University Press.

Farkas, J. 2020. A személyragozott és személyragozatlan főnévi igenév igeiségének mértéke. [The degree of verbalness of inflected and uninflected infinitives in Hungarian]. *Általános Nyelvészeti Tanulmányok* [Studies in General Linguistics] XXXII: 69–81.

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 GmbH, Internationaler Verlag der Wissenschaften.

Farkas, J., and G. Alberti. 2018. Complementation. In G. Alberti, and T. Laczkó (eds.), *Syntax of Hungarian: Nouns and noun phrases,* 662–775. Amsterdam, The Neatherlands: Amsterdam University Press.

Farkas, Judit – Veronika Szabó – Gábor Alberti 2017. Information-Structurally (Un)Ambiguous Nominal Constructions in Hungarian. *Finno-Ugric Languages and Linguistics* 6/2: 38–80.

Farkas, J., Szabó, V., and G. Alberti. 2021. Egy- és többértelműségek igéből képzett és közönséges főnevek csoportjaiban [(Un)ambiguity in phrases of deverbal and non-derived nominals]. *Általános Nyelvészeti Tanulmányok* [Studies in General Linguistics] XXXIII: 125–148.

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

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

Grohmann, K. K. 2000. Prolific Peripheries: A Radical Viewfrom the Left. Dissertaion at University of Maryland, USA.

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

Grohmann, K. K. 2009. Exploring features and arguments. In K. Grohmann (ed.), *Explorations of Phase Theory: Features and Arguments (Interface Explorations),* 1–20. Berlin, Germany: Mouton de Gruyter.

Horvath, J. 1997. The status of 'wh-expletives' and the partial wh-movement construction in Hungarian. *NL&LTh.* 15: 509–572.

Kenesei, I. 1998. Argumentumszerkezet és VP-fókusz. *A mai magyar nyelv leírásának újabb módszerei* III.: 223–242. Szeged: JATE.

Sigurðsson, H. A. 2009. Remarks on Features. In K. K. Grohmann (ed.), *Explorations of Phase Theory: Features and Arguments (Interface Explorations),* 21–52. Berlin, Germany: Mouton de Gruyter.

Szabolcsi, A. 2018. Strict and Non-strict Negative Concord in Hungarian: A Unified Analysis. In H. Bartos, M. den Dikken, Z. Bánréti, and T. Váradi (eds.), *Boundaries Crossed, at the Interfaces of Morphosytax, Phonology, Pragmatics and Semantics,* 227–242. Cham, Switzerland: Springer.

Varga, L. 2016. The Intonation of Topic and Comment in the Hungarian Declarative Sentence. *Finno-Ugric Languages and Linguistics* 5/2: 46–77.

LingBaW *Linguistics Beyond And Within 8 (2022), 75-85*

Postcolonial analysis of educational language policies of Ireland, Singapore, and Malaysia

Zuzanna Grala
Leipzig University, Germany

Abstract
The aim of this paper is to compare the educational language policing in Ireland, Singapore, and Malaysia. While distant geographically, the three countries experience similar linguistic processes when it comes to anglicisation, and propose different solutions to the issue of balancing linguistic rights, and promotion of English as the language of globalisation.

This comparison aimed to find out what influences language policing in postcolonial countries, and in what ways language shift can be prevented. The aspects of language policing strategies are presented as a way of protecting linguistic human rights, but also as a way of dealing with the aftermaths of the policies implemented by the British Empire.

Similarities and distinctions in the language policies of Ireland, Singapore, and Malaysia prove that the weak position of native languages originates not in the "natural" decline of a language, but rather in the policy of promoting English by the colonial forces. Ethnic and linguistic discrimination favouring English speakers in Ireland, Singapore, and Malaysia, originates in similar, imperial linguistic ideologies, which are still reflected in the current language policies of countries of colonial past. While the countries approach their bilingual educational policing in different ways, ultimately the outcomes seem similar when it comes to linguistic attitudes and prestige.

Keywords: language management, educational language policing, Ireland, Singapore, Malaysia

1. Introduction

The original aim of this paper was to analyse the attempts at the protection of the Irish language in the context of British influence and oppression. I believed that it could be beneficial to consider Irish language policing in a postcolonial context.

I was especially interested in comparing the linguistic issues of Ireland to other countries of colonial past. I chose Malaysia and Singapore as I hoped to identify elements of language policing that lead to different outcomes despite the countries' common past. My hope was to find out which aspects cause a country to become more anglicised and influence the prestige of other languages in the region.

The aim of this paper is to show that the ethnic and linguistic discrimination favouring English speakers in Ireland, Singapore, and Malaysia, originates in the imperial linguistic ideologies, which are still reflected in current educational language policies. While the countries approach bilingual educational policing in different ways, ultimately the attitudes towards commonly spoken languages show some similarities, especially when it comes to the role of English. The increased use of English, at the cost of other languages, is therefore not necessarily a "natural" occurrence, but a consequence of a continuing spread of imperial linguistic ideologies.

2. Background

Phillipson's theory of linguistic imperialism describes the power relations between English and other languages in postcolonial contexts, especially "the dominance of English […] asserted and maintained by the establishment and continuous reconstitution of structural and cultural inequalities between English and other languages" (Phillipson 1992: 47).

During the colonial period, the usual sociolinguistic situation of the colonised region was a binary opposition of the European language of the coloniser, used in formal contexts and specialized domains, and the local languages, used in local contexts only (Ricento 2009). In the Eurocentric, imperial worldview, English came to represent progress and prosperity (Phillipson 1992). It was also presented as a vehicle of "European" qualities, such as spirit and rationality, which in turn led to the establishment of a literary corpus, high culture institutions, history, and art (Wiley 2005). As a result of colonial influence, parts of the linguistic propaganda became assimilated into the cultures of (post-) colonial countries (Rassool 2013).

I find it beneficial to consider the outcomes of colonial language policing in the context of linguistic human rights, understood as a subfield of sociolinguistics that deals with the ways in which "social inequalities are legitimated and realised on the basis of language inequalities" (Schneider 2006: 17). Linguistic human rights operate on the basic assumptions that all human beings are equal, all languages are of equal value, and identification with and development of a person's mother tongue is a basic human right (Schneider 2006).

While the spread and popularity of English are often seen as a positive unifying element, Phillipson argues that, as an element and tool of imperialism, the English language spread contributes to the language shift processes ending in the death of other, indigenous, languages (Phillipson 1992). Educational policing is especially important in this context, as it plays a vital role in spreading the "ideology transmitted within and through the English language" (Phillipson 1992: 1). Despite the absence of a continued, direct colonial influence, linguistic imperialism contributes to linguistic attitudes and influences language policing of (post-) colonial countries.

3. Historical circumstances

The use of the Irish language has been restricted in various ways since the fourteenth century. The elimination of the Irish language and culture was one of the most vital aspects of the early colonisation or subjugation attempts of the English, and the ideologies of British and English superiority were already present in Ireland long before the administrative incorporation in 1800 (Crowley 2017). In the difficult times of the 19th century, English was seen as a way of escaping poverty and hunger, while Irish became associated with trauma and death (Crowley 2017). Despite that, the language continued to be seen as an important marker of national identity. Its role in the nationalist movement during the process of state formation resulted in a strong movement towards the preservation and protection of the language (Ó Croidheáin 2006). Since the beginning of the twentieth century, and as a consequence of the social and political circumstances, Irish language policing took a turn towards a less nationalist, bilingual approach, with English and Irish as official languages (Ó Croidheáin 2006).

The Malaysian Peninsula has been subject to European colonisation since the 16th century. The Straits Settlements became a British crown colony in 1889 (Schneider 2009). By that time, settler colonies were replaced by economic exploitation, which aimed to gain maximal profits by exploiting natural resources of the Peninsula and using Singapore a centre for trade between the Middle East, India, and China. This resulted in increased migration into the region, which in turn created a culturally and linguistically diverse population, where, by 1931, the Malay people were no longer a majority (Church 1997). The socioeconomic situation differed between and within groups, with the common element of the upper-class groups receiving English-medium education, which was to offer them more social and economic opportunities, and rural and poorer groups still using their original languages (Campbell 2018).

Even after the independence was gained in 1957, linguistic and ethnic issues did not disappear and eventually resulted in the separation of Singapore in 1965, and the policy of positive discrimination against the indigenous people (Albury and Aye 2016). A part of this policy was reducing the "emphasis on English education, which would favour the urban non-Malays" (Puteh 2010: 194). Singapore took another direction, promoting English medium education in an effort to "give a newly independent country a sense of belonging to a new nation and to differentiate it from the colonizer" (Suárez 2005: 462).

4. Current circumstances

In 2016, only 1.7% of the population of the Republic of Ireland spoke Irish every day outside the school system (Census 2016). Even in the protected Gaeltacht areas, Irish was used daily by only 32% of the population. The official language policy of the Republic of Ireland is focused on reviving the language as a vernacular and protecting the existing Irish-speaking communities. Unfortunately, despite various language revival policies, the numbers of Irish speakers are lower with every census. Currently, there are virtually no monolingual communities, even in the traditionally Irish regions in the west of Ireland. In most of the language policies, which take a careful bilingual approach, Irish is now presented as a

language of culture and identity, but not necessarily of economics or politics, which is a common outcome in colonised regions (Ricento 2009).

At the time of independence, Singapore struggled with a lack of a distinct national identity, separate from Malaysian or Chinese (Suárez 2005). As a result, English was chosen as the working language of the country, with Malay remaining a national language (Wee 2013). The national language was meant to connect Singapore to its Asian heritage, while English allowed it to compete globally and benefit economically (Suárez 2005). Currently, English, Malay, Mandarin, and Tamil are the four official languages of Singapore. Similarly to Ireland, Singapore adopts a policy of bilingualism, with students learning English and one of the other official languages at school (Wee 2011). In the context of the need to create a national identity separate from British or Chinese, the English variety Singlish, or Singaporean English, emerged as an identity marker for many Singaporeans. While fulfilling one of the aims of Singaporean language policing, the use of this variety is heavily criticised by the state, especially in official settings, as it is seen as a threat to the proficiency of English and its (perceived) many benefits (Murata and Jenkins 2009).

In Malaysia, as there are 14 big language groups and more than 80 smaller language communities, the aim of current language policing is to create linguistic unity through Bahasa Malaysia, which is the national language for all Malaysian people regardless of their ethnicity. The linguistic rights of other ethnic groups are restricted, in what Albury and Aye (2016: 71) call "limited acceptance of linguistic diversity". The policy of positive discrimination against ethnically Malaysian people can be quite harmful to other groups, beyond the issue of linguistic rights. The policy does not, however, pose a threat to the role of English. English is taught in schools as a foreign language, but it is important as a language of business, trade, and industry. This aspect made the effort to eliminate English from the public sphere in order to separate Malaysia from its colonial past more difficult, as English "continued to possess linguistic power and capital" (Gill 2005: 255).

5. Examples of educational language policing

5.1. Ireland and education as a tool for revitalisation

The hope for reviving the Irish language through teaching has been in place since the beginnings of the revitalisation efforts (Ó Croidheáin 2006). The educational goal after the Republic of Ireland gained independence was to prove that there are no significant disadvantages to bilingualism, and that Irish is not only suitable for the past (Ní Dhrisceoil 2013). Therefore, revitalisation of the language was understood as teaching the language to new generations, and not making sure that the language is continually used.

In the Republic of Ireland, education is the first out of 9 "areas for action" of the 20-Year Strategy for the Irish Language, a policy that aims to revitalise Irish as a community language ("20-Year Strategy" 2010). So far, this policy has been implemented relatively well. Therefore, it would be false to assume that the language has not been revived due to the lack of proper implementation of the educational language policy. Instead, it seems that, while the policies

have been implemented, they are not resulting in the desired outcome. Modern Irish language campaigns, such as the Bród Club, try to distance themselves from Irish school education, which is often associated negatively with the language (Kelly-Holmes and Atkinson 2017).

When it comes to the revitalisation efforts, one of the biggest problems is that, while in general there is an interest in promoting, learning, and speaking the Irish language, some people affected by the policy do not have any interest in it. Within the educational system, Irish is often deemed less relevant than other subjects. In their assessment of compulsory Irish examinations, surveyed students expressed the view that Irish is stopping them from learning what is more interesting, practical, or needed at university (Banks et al. 2018). Irish can also be seen as difficult, boring, or not useful (Duffy 2016). Such views, even if presented subjectively or anecdotally, reflect the attitude toward the Irish language and its pragmatic value outside the school system.

Another issue is that the policy is based on a false, or simplified, assumption, that language proficiency among children will result in proficiency among adults. The lack of daily language use creates a disconnect between Irish teaching and its revitalisation as a vehicular language. While the educational system succeeds in teaching children Irish and the percentage of Irish speakers is very high within the group of 5- to 16-year-olds, it drastically drops for people outside the school age (Census 2016). With only 2.6% of daily Irish input, a student who uses Irish only during Irish classes is unlikely to reach proficiency and confidence similar to that in English (De Barra 2019). As a result, Irish never becomes "natural" or "comfortable enough" to use when it is no longer mandated by the school. In a way, the Irish language policy lacks continuity, and there is no adequate support for Irish speakers of working age.

It seems that, in the case of a minoritized language, education policy is simply not enough to reverse the decline of daily language use. In general, while the Irish language policy provides support for already existing Irish-speaking communities in the Gaeltacht areas, and promotes the language to young people, there is no precise plan that would ensure continual use of the language. Even a relatively good educational language policy isn't enough to reverse the language shift. While language protection policies are in place, Irish-speaking communities are declining due to a lack of economic prospects, an aspect of language policing which is neglected in favour of a higher emphasis on education policing (Crowley 2017). Regardless of their preference, the economic, and social situation forces the speakers of a minority language to choose to function mainly in English, which they need in order to be employed in a particular sector, obtain higher education, or live in a city (May 2005). The Gaeltacht communities are getting older, and relying on Irish surviving in traditional Gaeltacht without providing support is simply unrealistic.

A more holistic approach could be beneficial when it comes to language revitalisation. Except for existing funding for language schools, and small businesses, the Irish-speaking areas, both traditional and those that emerged more recently, require better infrastructure, employment opportunities, and long-term planning when it comes to the development of the rural areas. This support, combined with a renewed interest in a rural way of life after the COVID-19 pandemic, creates an opportunity to close the gap between teaching Irish and its vernacular use.

5.2. Singapore and a pragmatic approach to linguistic education

In Singapore, the lack of intergenerational use of language is not of national concern. The education system is not aiming to revitalise minority languages or promote a specific regional language, and instead takes a more pragmatic approach, which would most likely be appealing to those with negative attitudes towards the teaching of Irish in Ireland. Singapore, which is only 687 sq km and does not have natural resources, relies on international trade and a global economy (Suárez 460). Therefore, the education policy includes English from a very early stage, and fluency is encouraged if not mandatory in hopes of facilitating the use of English towards "progress" and economic achievement (Schneider 2006).

At the turn of the century, Singapore was promoted as a country where one can benefit from the prestige and privileges that come with knowing English, and the language has kept a vital role in Singaporean society to this day. The linguistic policy of Singapore is based on an instrumentalist view of language planning, in which language is a tool, used to serve the economy or common values (Dixon 2009). The promotion of English as a language of business and economy, connected to both the imperial ideas of superiority of global English and the economic aspects of globalisation, contributed to the pragmatic decision to use it as a first language. As a result, the status of English is now stronger than those of the other three official languages of Singapore, and growing numbers of people speak it outside the business domain, as reflected in the latest census (Census 2020).

As the school system in Singapore is strongly separated between the ethnic groups, English-knowing bilingualism is especially valuable, as it is seen to enhance national integration, without privileging any of the three national languages (Gopinathan 1979). The main aim of the bilingual education policy is to make sure that the pupils achieve proficiency in English, which is supposed to offer them opportunities they do not have with their mother tongues (May 2016).

Culturally, the type of education that one receives is seen as a sign of intelligence, prestige, and influence (Dixon 2009). English-medium education is introduced at the earliest stages in hopes of achieving early proficiency. The languages are taught separately, without interreference, because of the view that the mother language does not contribute to second language acquisition. Because of the belief that the quantity of time devoted to studying a language is more important than quality, the language education takes up most of the student's school day and is even more extensive to those who are considered "skilled". While these aspects create a view of the Singaporean education policy as one that is focused on academic achievement and practical use of language, mother tongues are supposed to promote "Asian values" and "provide Singaporeans with a sense of ethnic identity" (Abu Bakar 2015: 46). The mother tongue, in its oral form, is to provide students with cultural and literary background characteristic to their ethnicity, rather than a means of non-vernacular communication (McKay and Bokhorts-Heng 2017). In this way, the mother tongues fulfil a similar role to that of Irish in Ireland and are taught similarly, together with culture, literature, and folklore (Dixon 2009). This is perhaps the main reason why the bilingual policy is maintained, despite the overwhelming difference in the prestige of the mother tongues and English.

Despite this effort to promote mother tongues along with English, the use of English is steadily growing. In the last ten years, the percentage of people who use English as a home language grew from 32 to 48%, mainly at the cost of Mandarin and other Chinese Dialects (Census 2020). Singaporean language policies are being adapted to the new reality: the Speak Mandarin Campaign expanded its goal to promote Mandarin not just among dialect speakers, but also English-educated and English-speaking Chinese Singaporeans ("About the campaign" 2022). The use of the Singaporean variety, Singlish, is partially disturbing this opposition between the cultural use of the Asian languages and the pragmatic, instrumental use of English (McKay and Bokhorts-Heng 2017). There is a chance that with the acceptance of Singlish as a marker of distinctively Singaporean, Asian identity, the importance of the three mother tongues will diminish. For the Singaporean linguistic situation, it would be a shift from polyglossia and multilingualism, to a diglossia with a few minority languages present. While this is still a hypothetical situation, it should be considered as a possibility within long-term language management schemes.

5.3. Malaysia and the conflicting goals of educational policing

As a result of the strong promotion of Malay culture, values, and language, the status of Bahasa Malaysia as a national language is indisputable. However, despite a strong nationalist approach to language policy, the goals of Malaysian language management are still conflicted when it comes to the education system. Educational language policy is influenced by two main factors: promotion of a coherent national identity, and striving towards a strong position of the country in the globalised world (Albury and Aye 2016).

The current state of the language education policies is a result of colonial policies, the linguistic plurality of Malaysia, and the early policies of using Bahasa Malaysia as the sole medium of education ("National Education System" 2015). Since Malaysia's independence, the educational language policies are conflicted between "local cultures and their demands", and "globalisation and internationalising aspects" (Puteh 2010: 195).

The Barnes Report from 1951, proposed by the British government, promoted a bilingual education policy, in which Malay would be the main language at the primary level, and English at the secondary level (Gill 2005) This policy combined the idea of uniting students of all ethnicities under Bahasa Malaysia and providing them with the opportunities which were perceived to come with English language education.

The following Razak Report (Report of the Education Committee 1956), promised "a place in primary school for every child and a unified educational system which promotes national unity and consciousness by using the national curriculum, not the national language" (Puteh 2010: 194). Razak Report supported not only Malay schools, but also vernacular schools and mother tongue education, which in colonial times were considered an "unreasonable [...] expenditure" (Gill 2005: 245). The new, independent education policy was dedicated to supporting disadvantaged Malay speakers and providing them with the same opportunities that were associated with English-speaking, urban areas (Albury and Aye 2016). Unfortunately, the implementation of Bahasa Malaysia policing was rather slow and ineffective, contributing to the culmination of the racial riots of 1969 (Gill 2005).

In the next twenty years, following the 1967 National Language Act, national schools and universities which used English as a medium of instruction were converted into Malay-medium institutions, and English was taught simply as a school subject (Azmi 2013). Between 1972 and 1988, there has been an extensive effort to adapt Bahasa Malaysia to its new role, which included the modernisation of the language, as well as the development of terminology (Gill 2005).

Just ten years after this conversion process was completed, economic aspects of language policing became more pressing, leading to an adjustment in the form of the Education Act 1996, seen by Puteh as a response to globalisation (Puteh 2010). The Education Act allowed English-medium education in some cases, provided that the national language was taught as a compulsory subject (Puteh 2010). At the same time, English was yet again allowed as a medium of instruction at private universities, which eventually lead to higher employment rates for their graduates (Gill 2005).

The next step towards globalisation and a pragmatic approach to language policing seen in Singapore was the Teaching and Learning of Science and Mathematics in English (PPSMI) policy, which as the name suggests, promoted English as the language of STEM subjects (Schneider 2009). While most reasons for this change were given as having to do with a lack of specialist vocabulary or access to scientific texts, the view of English providing students with a greater understanding of science seems to be rooted in the high prestige and perceived attributes of English, rather than in actual results of the students. The PPSMI policy was not particularly popular, as it was seen as a poor alternative to mother tongue education, which threatened other languages spoken in Malaysia, especially minority languages. After a period of intense protests and extensive corpus planning, the PPSMI policy was reversed in 2012, replaced by Upholding the Malay Language and Strengthening Command of English (Radhi 2020).

Just 8 years later, the prime minister of Malaysia announced that the government is reconsidering the reintroduction of the PPSMI policy (Soong 2020). This reversal would be especially influential for the Indian and Chinese students, as in the PPSMI policy the only language taught in their mother tongue would be the mother tongue itself, essentially diminishing the difference between Tamil/Mandarin-medium school and Bahasa Malaysia-medium school (Soong 2020). The decision is welcomed by some and dreaded by others, but the inconsistency in the Malaysian policy is criticised by both groups. Not only are the changes harmful to the students and teachers, but they also show the inconsistency and lack of long-term planning when it comes to Malaysian language policy. Even now, there is still a strong conviction that the native languages may be applicable to culture and home life, but to prosper in the field of science or economics, English is necessary, even on a national level. The current state of language education policies is therefore still an extension of the original, separated colonial policy, in which Bahasa Malaysia is sufficient for home life, but English is necessary to achieve globalisation and technological advancement (Schneider 2009).

6. Conclusion

Language policies of Ireland, Singapore and Malaysia are the result of the countries being faced with the choice between promoting English or the other language(s) of the nation. The outcomes reflect the conflict between a pragmatic choice of using the colonial language, and a more cultural, national approach (Puteh 2010).

Even without direct British influence, the three countries choose to promote English as the language of business, trade, and economy. Regardless of the national sentiment, English is seen as the language which offers opportunities better than those that come with the original language. The promoted model of language use, English-speaking bilingualism, reflects a divide between English as vehicular and other languages as gregarious languages of social intimacy, home life, and shared identity. In general, while the colonial aspects of Southeast Asian and Irish history are different, the issues of globalisation, linguistic imperialism, linguistic human rights and multilingual policing affect both similarly. The one common aspect which connects Ireland, Singapore, and Malaysia, is that all three countries strive to compete in the globalised economy, by the means of English. This aspect of the language beliefs, regardless of toleration-oriented rights of various linguistic communities, promotes the use of English as a means of success and gain, contributing to the decline of other languages.

The imperial linguistic ideologies of the language policies impact the attitudes towards all languages and contribute to ethnic and linguistic discrimination. While Malaysia is typically shown as being very distinct from Singapore, the changes in the last 50 years show a tendency to take a similar, less national, and more globalised approach to language policing. One may expect Irish policing to be more traditional, or independent of English, but the Anglicization of Ireland can be seen as even more advanced than that of Singapore. The processes that we see in Singapore, such as the development of a strong and distinct English variety, seem to already be completed, or at least more advanced, in Ireland. I hoped to compare the policies of Singapore and Malaysia to draw conclusions on how to best approach the Irish question. Instead, by analysing the history and current state of the Irish language, one can predict the future linguistic landscape of other postcolonial countries that take the instrumental, bilingual approach to language policing. Hopefully, taking a more balanced approach, that recognises the outcomes of linguistic imperialism and its impact on postcolonial countries, but also does not isolate the issue of English proficiency from that of language shift, language death, or the rights of linguistic minorities, can offer an opportunity to prevent language shifts before there is a need for language revitalisation.

References

20-year strategy for the Irish language 2010–2030: Progress report 2010–2015. Search for services or information. 2010. Retrieved January 23, 2022, from https://www.gov.ie/en/publication/9bf64f-20-year-strategy-for-the-irish-language-2010-2030-progress-report-20/

About the campaign. Who We Are. n.d. Retrieved November 28, 2022, from https://www.languagecouncils.sg/mandarin/en/about.

Abu Bakar, M. 2015. Malay, English and religion: Language maintenance in multilingual Singapore. *Issues in Language Studies* 4(1): 46–62. https://doi.org/10.33736/ils.1645.2015

Albury, N., and K. Aye. 2016. Malaysia's national language policy in international theoretical context. *Journal of Nusantara Studies (JONUS) 1*(1): 71–84. https://doi.org/10.24200/jonus.vol1iss1pp71-84

Azmi, M. N. L. 2013. National Language Policy and Its Impacts on Second Language Reading Culture. *Journal of International Education and Leadership 3*(1): 1–11.

Banks, J., McCoy, S., and E. Smyth. 2018. (working paper). *Senior Cycle Review: Analysis of discussions in schools on the purpose of senior cycle education in Ireland*. Economic and Social Research Institute. https://ncca.ie/media/3877/esri-cycle-1-analysis-nov-2018.pdf.

Campbell, J. 2018. English language and an inclusive Malaysia. *3L The Southeast Asian Journal of English Language Studies 24(*3): 206–219. https://doi.org/10.17576/3l-2018-2403-15

Casey, J. 2019, July 31. *Calls for 'change to current system' of teaching Irish in schools.* Irish Examiner. https://www.irishexaminer.com/breakingnews/ireland/calls-for-change-to-current-system-of-teaching-irish-in-schools-940764.html.

Census 2016 summary results – part 1 – CSO – central statistics office. CSO. 2018, October 4. http://www.cso.ie/en/csolatestnews/presspages/2017/census2016summaryresults-part1/

Census of demographic characteristics, 2020 ... – singstat. n.d. https://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1

Crowley, T. 2017. Language, Politics, and the State(s): Reflections from Ireland. *Harvard Ukrainian Studies 35*(1/4): 331–349. http://www.jstor.org/stable/44983547

Church, P. 1997. *Focus on Southeast Asia*. Allen & Unwin.

De Barra, C. 2019. *Opinion: 'you often hear that the Irish language is badly taught in our schools but that is not true'*. TheJournal.ie. https://www.thejournal.ie/readme/opinion-it-is-a-truth-universally-acknowledged-in-ireland-that-the-irish-language-is-taught-badly-in-our-schools-believe-it-or-not-this-isnt-true-4514137-Mar2019/

Dixon, L. Q. 2009. Assumptions behind Singapore's language-in-education policy: Implications for Language Planning and Second language acquisition. *Language Policy 8*(2): 117–137. https://doi.org/10.1007/s10993-009-9124-0

Duffy, R. 2016. Irish is considered among the hardest and least interesting subjects in school. *TheJournal.ie.* http://www.thejournal.ie/irish-language-school-2733847-Apr2016/

Gill, S. K. 2005. Language policy in Malaysia: Reversing direction. *Language Policy 4*(3): 241–260. https://doi.org/10.1007/s10993-005-7859-9

Gopinathan, S. 1979. Singapore´s language policies: Strategies for a plural society. *Southeast Asian Affairs 1979*(1): 280–295. https://doi.org/10.1355/seaa79v

Kelly-Holmes, H., and D. Atkinson. 2017. Perspectives on language sustainability in a performance era: Discourses, policies, and practices in a digital and social media campaign to revitalise Irish. *Open Linguistics* 3(1): 236–250. https://doi.org/10.1515/opli-2017-0012

May, S. 2005. Language Policy and Minority Rights. In T. Ricento (ed.), *An introduction to language policy: Theory and Method,* 255–272. Blackwell.

May, S. 2016. Globalization, Language(s), and Politics. In M. Gazzola, and B-A. Wickström (eds.), *The economics of Language Policy,* 383–402. The MIT Press.

McKay, S., and W. D. Bokhorst-Heng. 2017. *International English in its sociolinguistic contexts: Towards a socially sensitive Eil pedagogy*. Routledge.

Murata, K., and J. Jenkins. 2009. *Global Englishes in Asian contexts current and future debates*. Palgrave Macmillan.

National Education System-Brief History. History of The National Education System in Malaysia I The New Economic Model and Education. 2015. Retrieved January 23, 2022, from http://www.espact.com.my/national-education-system/brief-history

Ní Dhrisceoil, V. 2013. The Irish Language in Twenty-First-Century Ireland. In J. V. Ulin, H. Edwards, and S. T. O'Brien (eds.), *Race and immigration in the New Ireland,* 131–153. University of Notre Dame Press.

Ó Croidheáin C. 2006. *Language from below the Irish language, ideology and power in 20th-century Ireland*. Lang.

Phillipson, R. 1992. *Linguistic imperialism*. Oxford University Press.

Puteh, A. 2010. The language medium policy in Malaysia: A plural society model? *Review of European Studies 2*(2): 192–200. https://doi.org/10.5539/res.v2n2p192

Radhi, N. A. M. 2020, January 31. *Science and mathematics to be taught in English again: New straits times*. NST Online. Retrieved January 23, 2022, from http://www.nst.com.my/news/government-public-policy/2020/01/561363/science-and-mathematics-be-taught-english-again

Rassool, N. 2013. The Political Economy of English Language and Development: English vs. National and Local Languages in Developing Countries. In E. J. Erling, and P. Saergant (eds.), *English and development: Policy, pedagogy and globalization,* 45–68. Multilingual Matters.

Ricento, T. 2009. Language Policy: Theory and Practice- An Introduction. In T. Ricento (ed.), *An introduction to language policy: Theory and Method,* 10–24. Blackwell.

Schneider, B. 2006. *Linguistic human rights and migrant languages: A comparative analysis of migrant language education in Great Britain and Germany*. Peter Lang.

Schneider, E. W. 2009. *Postcolonial English: Varieties around the world.* Cambridge Univ. Press.

Soong, K. K. 2020. *New Education Reforms or Just Mahathir's Old Obsession?* FMT. Retrieved January 23, 2022, from https://www.freemalaysiatoday.com/category/opinion/2020/02/02/new-education-reforms-or-just-mahathirs-old-obsession/

Suárez, S. L. 2005. Does English rule? language instruction and economic strategies in Singapore, Ireland, and Puerto Rico. *Comparative Politics 37*(4): 459–478. https://doi.org/10.2307/20072904

Wee, L. 2011. *Language without rights*. Oxford University Press.

Wee, L. 2013. Language Policy in Singapore: Singlish, National Development and Globalisation. In E. J. Erling, and P. Seargant (eds.), *English and development. policy, pedagogy and globalization,* 204–219. Multilingual Matters.

Wiley, T. 2005. The Lessons of Historical Investigation: Implications for the Study for Language Policy and Planning. In T. Ricento (ed.), *An introduction to language policy: Theory and Method,* 135–152. John Wiley; Sons, Incorporated.

LingBaW *Linguistics Beyond And Within 8 (2022), 86-98*

Frames and political choice in Scottish election campaigns

Éva Jakusné Harnos and Mátyás Gergi-Horgos
University of Public Service, Budapest, Hungary

Abstract
The paper presents interdisciplinary research using the framework of cognitive linguistics based metaphor theory and nationalism studies of political science. Frames of movement are placed under scrutiny during the discourse analysis of the 2016 and 2021 election manifestos of the Scottish National Party and social media posts. In relation to metaphors of movement, images describing the future of an independent Scotland are also detected. The authors attempt to analyse and interpret findings both from the perspective of cognitive linguistics and ethno-symbolism. Apart from the texts of the manifestos, the timeframe of the research involved social media posts two months preceding and two months following the elections in both cases. Methodology was issue-driven and computer assisted but supervised: key words linked to movement were extracted from the manifestos and clustered. Their occurrence and frequency in the social media posts was checked. In the qualitative analysis phase, messages of the manifestos and of the posts were contrasted in order to answer our research question what kind of persuasive political discourse was used when options for the citizens were outlined.

Keywords: discourse analysis, cognitive metaphor, nationalism, elections, Scotland

1. Introduction

The Scottish elections held on 5 May 2016 resulted in the victory of the Scottish National Party (SNP) with 63 Members of the Scottish Parliament (MSP); two short of achieving a clear majority (Aiton et al. 2016: 3). The SNP secured its third electoral victory in the history of the devolved Scotland, being in power continuously since 2007; however, the elections also had another winner. The Scottish Conservatives increased their share of votes and became the largest opposition party in Holyrood, surpassing Scottish Labour (Anderson 2016: 559–560). Thus, Scottish party politics was now placed on a new spectrum where the voice of independent Scotland was still represented by the SNP, but for the first time, the nationalists' biggest pro-Union opposition group was not Labour, but the Scottish Conservatives. With the SNP in power and the Conservatives as leaders of the opposition in Scotland and government party in Westminster, the polarization over the constitutional question became even deeper, a

situation which was exacerbated further by the June 2016 referendum on the UK's EU membership (Simpkins 2017).

This new dynamism in party politics remained unchanged during the 2021 Scottish elections which took place on 6 May. Undoubtedly, the biggest winner of the day was the SNP, which, with 64 MSPs, secured more than double the number of seats won by the second placed party, the Scottish Conservatives (Aiton et al. 2021:3). The campaign of political parties was faced with the challenges of the COVID-19 pandemic, but relative optimism was also present as the rollout of vaccines was speeding up with 52 percent of the Scottish population receiving their first dose by 6 May (UK Government 2021). The results of the elections meant that the SNP came short of an overall parliamentary majority, but instead of forming another minority government, the nationalists made a cooperation agreement with the fourth largest parliamentary party, the Scottish Greens (Gilman and McKay 2021: 6–13). In the history of modern Scotland since the 1999 devolution, this was the third time a coalition government was formed to lead the country, but the first occasion that a pro-independence coalition was born with the prospect of holding another referendum on Scottish independence in the new parliamentary term.

2. Theoretical background: Nationalism, symbols and framing

Nationalism is a modern movement in politics, primarily concerned with elite-driven political and cultural transformation of a community (Breuilly 1993, 1996; Gellner 1983, Anderson 1983, Smith 2009). At the very core of the nationalist doctrine lies the central argument that a nation exists with distinctive character and the interests of the community must be prioritized over other considerations to ensure that the nation can secure independence (Breuilly 1993: 2). In this process, symbolic resources (traditions, memories, values, myths and symbols) play an important part as identification to specific community is not a static sense of belonging, rather a socially constructed and reconstructed attachment that is defined by political elites according to changing conditions and contexts (Özkirimli 2010: 88). Politicians give new meaning to already existing cultural symbols in a community, adapt old institutions and customs and even invent new traditions (Hobsbawm and Ranger 1983: 1–2). The reason for this is that these symbols can be used as political resources; they become referents of identification for the community, making it easier to forge political identity and use it in the struggle for political power (Brass 1991: 15; O'Leary 2001: 148). Cultural and symbolic elements of myth, memory, value, symbol and communication code thus provide frameworks for political aspiration, and therefore, the study of nationalism should be primarily concerned with analysis of these resources (Smith 1986: 14; Smith 2009: 18).

Our research used two theories of political communication as background: the functional theory of political campaigns (Benoit 2017) and political framing (Druckman 2010). The functional theory (Benoit 2017: 196–202) centres on campaign topics and the way candidates establish preferability when distinguishing themselves from opponents by acclaiming, attacking and defending. Campaign discourse on their personality and programme is disseminated by various sources. As to candidates' character, personal qualities, leadership ability and ideals are presented. After assessing policy options, constituents are encouraged to

vote either retrospectively, based on a candidate's past deeds and merits, or prospectively, based on a means–end analysis of future plans and goals.

Framing has multiple definitions dependent on the discipline (e.g. psychology, linguistics, media studies) where it appears. Cognitive frames or conceptual frames are semantic domains which relate to particular events, each containing certain scripts, which comprise particular roles (Goffman 1974; Schank and Abelson 1977; Brown and Yule 1991; Ziem and Schwerin 2014). Framing is closely related to context in broad sense. The constituents of context are situation and language, whose perception by the participants in communication can be negotiated, consequently, it is considered dynamic (Verschueren 1999: 87–95). Out of the three decisive factors of the situation (the mental, the physical and the social world), the mental world of the participants in a communicative situation, that is, the cognitive and emotional condition of a targeted individual or community is the easiest to impact. This explains why framing is exploited in political communication: certain aspects of a topic may be foregrounded while others may seem non-existent or irrelevant. Thus, in media framing, for instance, an episodic news frame (a concrete, event- or person-related news report) or a thematic news frame (a more abstract news report on policy) can be selected. The significance of framing for political communication is that it prepares the target audience for recognising and complementing scripts and frames even when they are not explicitly worded. As a result, voters are self-persuaded as they believe they have developed a perspective for themselves. According to Druckman (2010: 187), in campaign communication framing may range from 'equivalence framing' (attributing minimal difference to two things) to 'emphasis framing' (placing something in two opposing contexts suggesting contradictory judgement), the latter being more typical in competing political programmes. Obviously, framing in political communication is intended to influence the voters' attitude and behaviour.

Metaphors and other figurative language are efficient tools of shifting frames nearly unnoticed, in other words, of re-contextualising facts and events. Cognitive linguistics theories have proven that metaphors and other figurative language are not simply rhetorical patterns, rather, ways of conceptualising things of both physical and social reality, including politics and international relations (Lakoff and Johnson 1980; Lakoff 1996; Gibbs and Steen 1999; Handl and Schmid 2011; Gibbs 2017). Cognitive linguistics research has produced a great amount of literature on the role of figurative language in discourse (for example, Gibbs 2008; Halverson and Engene 2010; Kövecses 2010; Semino and Demjén 2017). Political discourse analysis has integrated metaphor and metonymy as means of examination of political concepts (Carver and Jernej 2008; Rycker and Don 2013; Wodak and Frochtner 2018). Consequently, cognitive metaphor theory seems to be applicable in research into political framing as an impactful means of persuasion.

An attempt to interpret metaphors in the context of ethnosymbolism may prove fruitful. Since metaphors are closely interrelated with culture, we assumed ties between metaphors used in the SNP campaigns and Scottish cultural traits. As Anthony D. Smith stated, 'Landscape, language, ethno-history, public religion and rituals are foundations of ethnic traditions and national identity, which may provide cohesion to a national state' (2007: 335). In our analysis an investigation of metaphors of movement seemed appropriate for understanding the justification of Scottish independence efforts.

3. Methodology

A comparative quantitative and qualitative analysis of the 2016 and of the 2021 Scottish National Party Manifestos was conducted. We applied computer-assisted methods for a corpus-based analysis with a supervised approach (Franzosi 2018; Kutter 2018), using AntConc Tool. Regarding the political framing of independence efforts, we focused on conceptual metaphors which can nearly unnoticeably influence voters because they are deeply embedded in culture. Our focus was metaphors of movement representing political decisions and political action.

For identifying metaphors of movement referring to progression or the lack of it, we used the methodology of the Pragglejaz Group (Pragglejaz Group 2007; Steen 2007). After reading the manifestos several times, we double-coded verbs of movement for literal and non-literal meaning in context and grouped the ones found metaphoric into three categories: progress, stagnation/stationarity and regression. Then a quantitative analysis of the two manifestos was conducted with AntConc. Our presumptions were as follows:

Hypothesis 1 was that the change independence would mean would be described by the conceptual metaphor of progress (movement ahead), whereas failing to become independent would be equalled with stagnation or regression.

Hypothesis 2 was that references to Scottish identity would be made by citing shared heritage such as customs, language, historical monuments. In order to test this presumption, we have included in our research key word references to Scotland and Scottish people as well as allusions to cultural peculiarities.

All the language tools are considered functional in the campaigns, that is, they serve the purpose of persuasion. This is especially true for metaphors and metonyms as suggested by available literature (Charteris-Black 2011; Musolff 2015; 2016;). A limitation of our research is, however, that we have selected and focused only on the possible link between metaphors of movement and political choice.

4. General remarks on the brochures

The title of the 2016 brochure is *Re-elect* and the digital version of the brochure proposes downloading a vision app for further content. Its print copies were disseminated by post. Its cover bears the photo of Nicola Sturgeon and the first pages (2016: 1–3) provide a visual summary of the major achievements of the previous term of the SNP government. The section *Stronger for Scotland* describes the performance of the incumbent government in safeguarding national interests (2016: 4–5). Pages 6 and 7 include a photo of Nicola Sturgeon and her short, print message, page 8 and 9 highlight the major points of the party programme. An interesting feature of the 2016 brochure is the short, thematic messages by supporters of the SNP coming from various social backgrounds, which are illustrated by attractive photographs probably taken in their homes or everyday environment (2016: 10; 14; 18; 20; 22; 24). These show similarities to episodic news frames discussed above. The reader gets the impression they will read a programme compiled by "ordinary people", which makes them more receptive to the messages of the text of the manifesto. Only then begins the presentation of the formal text of

the programme with the contents page. It will be discussed in more detail below, in the section titled Framing independence endeavours in 2016.

The title of the 2021 brochure is *Scotland's Future*, which indicates that the future vision of an independent Scotland is placed in the focus. Nicola Sturgeon's photos are on the cover and on page 3, with her address between. The brochure has a digital version, which we have used for analysis, but it was also disseminated in print form. Remarkably, the digital version is accessible to all people having a disability. Its structure is discussed below in the section titled Framing independence endeavours in 2021. The graphics depict the people it is intended for, notably including faces from diverse age groups and ethnic groups, visually reinforcing the content on creating a modern, multi-cultural and multi-lingual, independent Scotland. Episodic news frames are applied about ordinary citizens who support the SNP. With short video messages by leaders of the SNP and supporters (2021: 14–15) it gives the impression that the reader can hear "the people's voice". On page 44, small clickable thematic icons guide the user to further details of the key programme points of the SNP. On page 75, a short video can be clicked on and launched, in which a young female voter summarizes the events of the previous year of the COVID pandemic, also describing the grievances of the Scottish people caused by the UK government and encouraging citizens to vote for SNP and independence.

In summary, the title of the 2016 brochure implies that achievements of the government and of the First Minister are centred on, probably encouraging retrospective voting, while the 2021 brochure seems to be more focused on the future of an independent Scotland, suggesting a strategy to gain prospective votes.

5. Findings and discussion

Hypothesis 1

An underlying feature of both the 2016 and 2021 manifestos was an event-structure metaphor (Lakoff 1993; Kövecses 2010), which could be summarized as a construction from the conventional metaphors PROGRESS IS MOVING FORWARD; CHANGES ARE MOVEMENTS; LONG-TERM, PURPOSEFUL ACTIVITIES ARE JOURNEYS; PURPOSES ARE DESTINATIONS (Kövecses 2010:162–165). This is reflected by the data on frequency of the elements in the categories. The tables below present data on the frequency of verbs of progression, verbs of stationary and verbs of regression in the 2016 manifesto entitled *Re-elect*.

Table 1: *Most used verbs of progression in the 2016 brochure*

Categories of verbs	Occurrences
continue	116
bring; bring forward	39
progress; make progress;	18
increase, deliver, accelerate, see, sustain, support progress	
act; take action; reinforce action	10
take steps	9
leave	3
go further	2
shift	2

replace	1
advance	1
Total number of most used verbs of progression:	**208**

Table 2: *Most used verbs of stationary in the 2016 brochure*

Categories of verbs	**Occurrences**
maintain	26
keep	13
remain	13
stand	11
stop; halt	5
stay	5
Total number of most used words of stationary	**73**

Table 3: *Most used verbs of regression in the 2016 brochure*

Categories of verbs	**Occurrences**
return	2
fall behind	1
go back	1
refuse to take steps	0
send into decline	0
does not take action	0
take wrong direction	0
Total number of most used verbs of regression	**4**

The most prominent feature of the data from 2016 is that references to movement ahead substantially exceed references to stationarity and, especially, regression. Another noticeable fact is that movement ahead is identified with independence (the word occurs 7 times, as well as "independent" in the 2016 manifesto). For instance, the section entitled Moving Scotland Forward starts with the following:

(1) *We will achieve independence only when the majority of our fellow citizens are persuaded that it offers the best future for our country* (2016:23).

Thus, the efforts for independence are described metaphorically as progressing on a journey towards a destination which is the objective of political action: an independent Scotland. On the contrary, abandoning the objective of independence is described metaphorically as stopping along the journey and objections to attaining the goal of independence are impediments blocking the way forward.

(2) *At the same time if there is a clear demand for a referendum no politician has the right to stand in the way of the people of Scotland to choose their own future* (2016:24)

In comparison to the data in the tables above, statistics on the same categories proved to be different in the 2021 manifesto, as it can be seen below.

Table 4: *Most used verbs of progression in the 2021 brochure*

Categories of verbs	Occurrences
continue	52
bring; bring forward	32
move; remove	24
progress; make progress;	22
increase, deliver, accelerate, see, sustain, support progress	
replace	7
take steps	7
shift	5
leave	5
leave	5
act; take action; reinforce action	5
advance	2
Total number of most used verbs of progression:	**174**

Table 5: *Most used verbs of stationary in the 2021 brochure*

Categories of verbs	Occurrences
remain	31
keep	23
maintain	18
stop; halt	6
stay	4
stand	2
Total number of most used words of stationary	**84**

Table 6: *Most used verbs of regression in the 2021 brochure*

Categories of verbs	Occurrences
return	6
take wrong direction	2
go back	2
refuse to take steps	1
send into decline	1
does not take action	1
fall behind	1
Total number of most used verbs of regression	**14**

One explanation for the significant increase in references to stationarity and regression is that the SNP got strong rival parties, the Scottish Conservatives, and the Scottish Greens, in the election campaign, which required clearer alternative visions of the future. In addition, Brexit reinforced the future vision of an independent Scotland and served as a strong argument for fighting for it in the hope of firm support from the European Union into which it may return. Remarkably, the tone of the 2021 manifesto is more determined. For example, it contains 18 mentions of *independent* and 16 of *independence*, which is more than double of the 2016 occurrences. More importantly, outspoken criticism and political attack are targeted at the Westminster government by the 2021 manifesto.

The metaphoric presentation of the complex political procedure of Scotland becoming independent can be summarised in Figure 1 below.

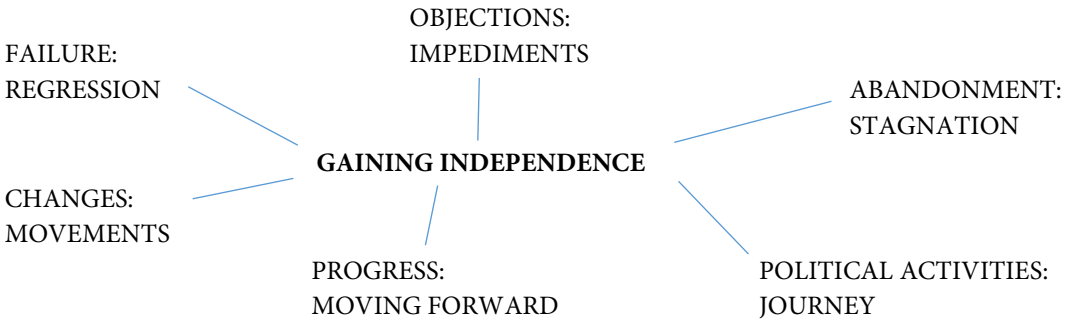


Figure 1: *The event-structure metaphor of gaining independence based on Kövecses (2010: 163)*

Hypothesis 2

We presumed that both manifestos would include allusions to Scottish identity and culture. Nevertheless, references to Scotland, Scots and other related words were diverse in amount.

Table 7: *The frequency of references to Scotland and its people in the 2016 and 2021 brochures*

Word	Occurrences in 2016	Occurrences in 2021
Scotland	318	361
Scottish	170	107
People	144	178
National	121	97
Communities	78	64
Community	62	52
Nation	19	16
Gaelic	4	18
Scots	2	7

It appears that *Scotland* was used both in its literal meaning (the geographical area) and non-literal (the state and its population) more increasingly in 2021, but mentions of *Scottish* and *nation* decreased in number. In contrast, references to Gaelic more than quadrupled by 2021. *Nation* and *national* seem to be slightly avoided in 2021, alongside with *community* and *communities*, which might be deemed too ethnicity-centred or nationalist. This may be underpinned by the increased use of *people*. *People* may be judged a more inclusive and stylistically less loaded term than *nation* or even *community*. The quote below is an example for innovative framing of the revival of national language and preservation of national culture.

(3) *We will also bring forward a new Scottish Languages Bill which takes further steps to support Gaelic, acts on the Scots language and* ***recognises that Scotland is a multilingual society*** (2021:66). (Emphasis added by the authors.)

This illustrates how love for one's ancient native language is conceptualised in terms of globalisation and modern political discourse, and thus re-interpreted and re-framed. The quote implies that the Scottish National Party is tolerant and inclusive and is in harmony with multi-culturalism and multi-lingualism. This trend fits well into the Scottish political elite's

project of creating a so-called "multicultural Scotland" which sees the country as a civic and inclusive nation that wants to join the "family of nations" through independence (Leith and Soule 2012; El Fekih Said 2018; Keating and McGarry 2001; Bechhofer and McCrone 2009).

Framing independence endeavours in 2016

The contents page of the 2016 manifesto includes the following themes: A Healthier Scotland; A Smarter Scotland; A Wealthier Scotland; A Fairer Scotland; Scotland's Future; A Thriving Rural Scotland; A Greener Scotland; An Empowered Scotland; A Safer Scotland; An International Scotland; A Creative Scotland. The comparative forms of adjectives imply that the vision of the future involves improvement and development compared with the current state of the country, which is linked to independence. It is summarised as follows:

(4) *We will undertake new work, starting this summer, with the aim of persuading a clear majority of people in Scotland that independence is the best future for our country.* (2016:11)

The language of the manifesto uses key words of the modern global and EU discourse, such as *innovation* (21), *internationalisation* (21), *sustainable* (27). On the whole, emphasis is laid on the description of Scottish independence efforts as fitting the modern global world:

(5) *Scotland will use its international engagement to continue to act as a good global citizen.* (2016:41)

Unique Scottish tradition is mentioned in the section Thriving Rural Scotland, below the subheading Modernising Crofting (2016:25), when the community ownership of land known from national history is discussed. The sub-section includes a promise to maintain croft housing and establish a new system of new woodland crofts, apart from modernising legislation on crofting. This could be an example for filling an old, historical concept in the shared background knowledge of a community with new meaning. Crofting is actually framed as suiting the modern goal of sustainable agricultural production.

Hutting, which is mentioned on page 27, was another peculiar characteristic of Scottish lifestyle between the two world wars, when industrial workers used to rent a small plot of land on which they built a small hut from timber. These huts had no running water and electricity in the past, which counted harsh conditions, however, currently it is regarded eco-conscious and eco-friendly (*What is Hutting?* n.d.). The manifesto promises to promote and help maintain the historically rooted hutting lifestyle. It is another example for a deep-rooted tradition reconceptualised and revived in the framework of modern life. Beside these two peculiarities of Scottish culture, the Gaelic language and Gaelic Art (2016:43) are mentioned, together with a Historic Environment Scotland Strategy (2016:44). On the whole, old Scottish lifestyle and independence are presented as the natural way of life of the people living in Scotland. Contrary to our expectations founded on ethno-symbolism, their Scottish nature is less emphatic, probably because of the impact of global culture and the objective of creating an inclusive, multi-cultural society.

Framing independence endeavours in 2021

The 2021 manifesto is structured around the following themes according to the contents page: Leading Scotland out of the COVID crisis; A democratic nation; Our vision; A caring society; An equal society; Living better; An economy that works; Land of opportunity; A net zero nation; Scotland in the world. As suggested by the section titles, the independence ambitions of Scotland are placed within the conceptual frames of a fair and inclusive society striving for the equality of citizens; the benefits of global economy, and protection of the environment, especially preventing climate change.

The language used for presenting the ideas integrates key phrases of modern EU discourse and globalist discourse. For example:

(6) *Scotland is a **welcoming and inclusive nation**.* (2021:72)
(7) *We **rely on migration** to grow our population.* (2021:72)
(8) *We will work **to build Scotland's population sustainably**.* (2021:72)
(9) (*Over the course of the next parliament we will:*) *Create a new global affairs framework, underpinned by Scotland's **fundamental values and priorities** and adopt a **feminist foreign policy**.* (2021:72)
(10) *Support the establishment of an **Institute for Peacekeeping** and a **Scottish Council for Global Affairs**.* (2021:72)
(11) *We place great importance on Scotland being **a good global citizen**.* (2021:74)

(emphasis added by the authors)

The examples prove that nationalist ambitions are framed as up-to-date and conforming to the trends in international relations. At the same time, the brochure refutes possible criticism claiming that independence would result in isolation and lack of development. However, whereas in 2016 certain cultural elements and Scottish lifestyles acted as references of independence endeavours, these were completely missing in 2021. The possible explanation for this is that the nation-building project has 'progressed' even further towards an outward looking and multicultural nationalism. Cultural references were left behind or were re-framed (such as Gaelic being understood as one of the many languages of Scotland) to fit the image of an inclusive and welcoming community. Therefore, a rather interesting, but seemingly contradictory pro-independence discourse emerged which defines the independent Scottish nation in terms of a globalist conceptual framing.

6. Conclusion

In our research metaphors of movement were examined through quantitative and qualitative analysis. A quantitative and qualitative analysis of key words was also conducted.

The independence efforts of the Scottish National Party are reflected in their 2016 and 2021 election campaign manifestos, in which an event-structure metaphor describing the complex process of gaining independence can be detected. Political struggle for gaining independence is conceptualised as movement forward during a journey, and problems are presented as blocking the way. Independence is progress, which suggests that the means and the end-state are interchanged. Abandoning the endeavour for independence would mean

regression or stationarity. Consequently, independence is described as a necessity for making progress. Hypothesis 1 has been supported.

The key words analysis of references to Scotland and its people did not confirm our expectations that a strong link could be established between allusions to cultural heritage and persuasive political communication. Instead, functional discourse was developed from a combination of SNP-related, EU-related and global discourse, which served the purpose of acceptance of the independence efforts home and abroad. Although it exploited a few culturally loaded words and concepts (especially in the 2016 manifesto), it did not fully underpin Hypothesis 2 concerning the cohesive force of shared cultural symbols and background knowledge of an ethnic group.

Our research findings are limited to the metaphor cluster of progress, stationarity and regression. Further examination of the conceptual metaphors may offer an explanation to the communication technique which has successfully mingled discourse about an independent Scotland, discourse about current global issues, and key words from shared Scottish background knowledge. Research should be extended to other election campaign brochures in order to gain a better understanding of the development of the SNP discourse.

References

Aiton, A., Burnside, R., Campbell, A., Edwards, T., Liddell, G., McIver, I., and A. McQuillen. 2016. *SPICe Briefing: Election 2016*. Edinburgh: Scottish Parliament. Available: https://archive2021.parliament.scot/ResearchBriefingsAndFactsheets/S5/SB_16-34_Election_2016.pdf (Accessed 10 October 2021).

Aiton, A., Atherton, S., Burnside, R., Campbell, A., Hudson, N., McIver, I., and Robinson, E. 2021. *SPICe Briefing: Election 2021*. Edinburgh: Scottish Parliament. Available: https://sp-bpr-en-prod-cdnep.azureedge.net/published/2021/5/11/591dc3c7-d994-4bbd-8120-767e9e781a67/SB%2021-24.pdf (Accessed 10 October 2021).

Anderson, B. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Bechhofer, F., and D. McCrone (eds.) 2009. *National Identity, Nationalism and Constitutional Change*. London: Palgrave Macmillan.

Benoit, W. L. 2017. The Functional Theory of Political Campaigns. In K. Kenski, and K. Hall Jamieson (eds.), *The Oxford Handbook of Political Communication,* 196–204. New York: Oxford University Press.

Brass, Paul R. 1991. *Ethnicity and Nationalism: Theory and Comparison*. New Delhi: SAGE Publications.

Breuilly, J. 1993. *Nationalism and The State*. Manchaster: Manchaster University Press.

Breuilly, J. 1996. Approaches to Nationalism. In G. Balakrishnan (ed.) *Mapping the Nation,* 146–174. London: Verso.

Brown, G., and G. Yule. 1991. *Discourse Analysis*. Cambridge: Cambridge University Press.

Carver, T., and P. Jernej (eds.) 2008. *Political Language and Metaphor*. Routledge, Taylor and Francis Group.

Charteris-Black, J. 2011. *Politicians and Rhetoric.* Palgrave Macmillan.

Druckman, J. N. 2010. Competing Frames in a Political Campaign. In B. F. Schaffner, and P. J. Sellers (eds.), *Winning with Words. The Origins & Impact of Political Framing,* 101–120. New York/London: Routledge, Taylor & Francis Group.

El Fekih Said, W. 2018. Reconstruction and Multiculturalism in the Scottish Nation-Building Project. *Études écossaises*, (20), DOI: https://doi.org/10.4000/etudesecossaises.1418.

Franzosi, R. 2018. Content analysis. In R. Wodak, and B. Frochtner (eds.), *The Routledge Handbook of Language and Politics,* 153–168, London/New York: Routledge, Taylor and Francis Group.

Gellner, E. 1983. *Nations and nationalism*. Ithaca: Cornell University Press.

Gibbs, R. W., and G. J. Steen (eds.) (1999). *Metaphor in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gibbs, R. W. (ed.) 2008. *Metaphor and Thought*. Cambridge: Cambridge University Press.

Gibbs, R. W. Jr. 2017. *Metaphor wars: Conceptual metaphors in human life*. Cambridge: Cambridge University Press.

Gilman, L., and S. McKay. 2021. *Scottish Government-Green Cooperation Agreement and other forms of political agreement*. Edinburgh: Scottish Parliament. Available: https://sp-bpr-en-prod-cdnep.azureedge.net/published/2021/9/29/951cd002-5c86-4f58-912d-09652cdc6cea/SB%2021-66.pdf (Accessed 10 October 2021).

Goffman, E. 1974. *Frame Analysis: An Essay on the Organization of Experience*. London: Harper and Row.

Halverson, S. L., and J. O. Engene. 2010. Domains and Dimensions in Metonymy: A Corpus-Based Study of Schengen and Maastricht. *Metaphor and Symbol* 25(1): 1–18, DOI: 10.1080/10926480903538456

Handl, S., and H.-J. Schmid (eds.) 2011. *Windows to the Mind. Metaphor, Metonymy and Conceptual Blending*. Berlin/New York: De Gruyter Mouton.

Hobsbawm, E., and T. Ranger (eds.) 1983. *The Invention of Tradition*. Cambridge: Cambridge University Press.

Iyengar, S. 2010. Framing Research. The Next Steps. In B. F. Schaffner, and P. J. Sellers (eds.), *Winning with Words. The Origins & Impact of Political Framing,* 185–191. New York/London: Routledge, Taylor & Francis Group.

Keating, M., and J. McGarry. 2001. *Minority nationalism and the changing international order*. New York: Oxford University Press.

Kövecses, Z. 2010. *Metaphor. A Practical Introduction*. New York: Oxford University Press.

Kutter, A. 2018. Corpus analysis. In R. Wodak, and B. Frochtner (eds.), *The Routledge Handbook of Language and Politics,* 169–186. London/New York: Routledge, Taylor and Francis Group.

Lakoff, G., and M. Johnson. 1980. *Metaphors We Live By*. Chicago/London: The University of Chicago Press.

Lakoff, G. 1993. The Contemporary Theory of Metaphor. In A. Ortony (ed.), *Metaphor and Thought,* 202–251. Cambridge University Press.

Lakoff, G. 1996. *Moral Politics: What Conservatives Know What Liberals Don't*. Chicago: University of Chicago Press.

Musolff, A. 2015. *Metaphor and Political Discourse. Analogical Reasoning in Debates About Europe*. Palgrave Macmillan.

Musolff, A. 2016. *Political Metaphor Analysis*. Bloomsbury Publishing.

O'Leary, B. 2001. Instrumentalist Theories of Nationalism. In. A. S Leoussi (ed.), *Encyclopedia of Nationalism,* 148–153. New Brunswick and London: Transaction Publishers.

Özkirimli, U. 2010. *Theories of Nationalism: a Critical Introduction*. Hampshire: Palgrave Macmillan.

Pragglejaz Group 2007. MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol* 22(1): 1–39.

Rycker, A. D., and Z. M. Don (eds.) 2013 *Discourse and Crisis. Critical Perspectives*. Amsterdam/Philadelphia: John Benjamins Publishing Co.

Schank, R.C., and R.P. Abelson. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Erlbaum.

Semino, E., and Zs. Demjén (eds.) 2017 *The Routledge Handbook of Metaphor and Language*. London/New York: Routledge, Taylor and Francis Group.

Simpkins, F. 2017. The 2016 Scottish Parliament Elections: Unionist Parties and the Constitutional Divide. In *Revue française de civilisation britannique* 22: 4, DOI: https://doi.org/10.4000/rfcb.1576.

Smith, A. D. 1986. *The Ethnic Origins of Nations*. Oxford: Blackwell Publishing.

Smith, A. D. 2007. Epilogue: The Power of Ethnic Traditions in the Modern World In A. S. Leoussi, and S. Grosby (eds.), *Nationalism and Ethnosymbolism. History, Culture and Ethnicity in the Formation of Nations,* 325–336. Edinburgh: Edinburgh University Press.

Smith, A. D. 2009. *Ethno-Symbolism and Nationalism: A Cultural Approach*. London: Routledge.

Steen, G. 2007. Finding Metaphor in Discourse: Pragglejazz and Beyond. *Culture, Language and Representation* V: 9–25.

Stewart Leith, M., and D. Soule. 2011. *Political Discourse and National Identity in Scotland*. Edinburgh: Edinburgh University Press.

UK Government 2021. *Vaccinations in Scotland*. Coronavirus.data.gov.uk. Available: https://coronavirus.data.gov.uk/details/vaccinations (Accessed 10 October 2021).

Verschueren, J. 1999. *Understanding Pragmatics.* London/New York: Arnold.

What is hutting? Exploring the new Scottish lifestyle trend. n.d. Available: https://www.lovecampers.co.uk/blog/conscious-campers/what-is-hutting-exploring-the-scottish-lifestyle-trend/ (Accessed 30 January, 2022).

Wodak, R., and B. Frochtner (eds.) 2018. *The Routledge Handbook of Language and Politics.* London and New York: Routledge, Taylor and Francis Group.

Ziem, A., and C. Schwerin. 2014. *Frames of Understanding in Text and Discourse*. Amsterdam: John Benjamins Publishing Company.

LingBaW *Linguistics Beyond And Within 8 (2022), 99-113*

Scottish Gaelic political terminology – Term formation in the Scottish Parliament Annual Report

Lena Krochmann
Leipzig University, Germany

Abstract
This paper analyses the Scottish Gaelic political terminology used by the Scottish Parliament, based on the translations of three Annual Reports. Applying principles from terminology theory and taking minority-language concerns into account, the Scottish Gaelic terms are examined for methods of term formation and their relation to their English and, where applicable, Irish equivalents. Inconsistencies in the terminology are also considered. The results are placed in the wider context of research on Scottish Gaelic terminology and corpus planning. The paper argues that the examined terminology is shaped by the dominance of English, as well as by practical concerns arising from the current language planning situation.

Keywords: Scottish Gaelic, terminology, corpus planning, translation

1. Introduction

After a long period of exclusion from most of public life and politics, Scottish Gaelic has become more present in these areas over the past few decades. Notably, the new Scottish Parliament, which opened in 1999, allows for the use of the language. In practice, however, this is mostly limited to the translation of selected publications. Scottish Gaelic also gained legal recognition under the Gaelic Language (Scotland) Act 2005, which refers to it as "[…] an official language of Scotland commanding equal respect to the English language […]". Promoting the use of Scottish Gaelic in all areas of life is one of the aims of official language policy, as described in the National Gaelic Language Plans (Bòrd na Gàidhlig 2007, 2012, 2018). This expansion of Scottish Gaelic use requires corpus and terminology planning. As there is no language academy or similar central corpus planning institution, several organisations are involved in the creation of specialised terminology. For instance, the Scottish Parliament commissioned two parliamentary dictionaries (*Faclair na Pàrlamaid* 2001 and *Faclair Rianachd Phoblaich* 2012), which involved the creation of new terms. The decentralised approach to Scottish Gaelic corpus planning has been criticised for a lack of co-

ordination and consistency (cf. Bauer et al. 2009; Dunbar 2010; Bell et al. 2014). Weaknesses of official Scottish Gaelic terminology have also been pointed out by McLeod (2000, 2001).

This paper, based on research from an unpublished MA thesis (Krochmann 2019) takes a closer look at one example of Scottish Gaelic specialised terminology: the political terminology currently used by the Scottish Parliament. Based on three of the Parliament's translated Annual Reports, it analyses the term formation in an attempt to characterise the Parliament's current Scottish Gaelic political terminology. The paper considers the different term types found in the reports and assesses the tendencies and preferences that can be inferred from them. It also discusses how the Scottish Gaelic terminology relates to its English counterpart, as well as political terms in Irish. The results can be interpreted against the backdrop of the current sociolinguistic and language planning situation for Scottish Gaelic. It is also useful to compare them to earlier research to assess how tendencies in Scottish Gaelic political terminology have evolved over time.

2. Background and methodology

Although the Parliament makes provisions for Scottish Gaelic, it is by no means a completely bilingual institution. There is the post of Gaelic Development Officer for various Gaelic-related tasks, but no in-house translation department.[1] English remains the sole language of legislation, and only some of the Parliament's official publications are translated. As a result, the amount of Scottish Gaelic text available from the Parliament is limited. As a regular bilingual publication covering a wide range of political topics, the Scottish Parliament Annual Report (which has been available in Scottish Gaelic since 2012) seems particularly suited for a terminology analysis. The text corpus for this analysis comprises the reports for 2016–17, 2017–18 and 2018–19, which were published in bilingual format (English and Scottish Gaelic translation). The word count of each bilingual report, as well as that of the translation alone, is given below.

***Table 1**: Word count of the text corpus*

Year	Words in bilingual report	Words in Scottish Gaelic translation only
2016–2017[2]	11,088	5,660
2017–2018	10,709	5,674
2018–2019	10,386	5,456
	32,183	16,790

When discussing political terminology, it should be noted that the research object is difficult to define. There are varying views on what is political and, accordingly, what constitutes political language and vocabulary. As this paper is concerned with terminology planning and the vocabulary used by a specific state institution, the focus is on the specialist terminology

[1] I would like to thank Alasdair MacCaluim, Gaelic Development Officer at the Scottish Parliament, for the information about translation and the use of Gaelic in the Scottish Parliament.

[2] This report contains a list of all members of the Scottish Parliament. As only the English versions of the names are given, this list is included in the overall word count, but not the translation word count.

used to describe a system of government and the associated processes. This type of political terminology is described by Klein (1989) as institution vocabulary (*Institutionsvokabular*) and comprises terms for state institutions and organisations, roles and offices, laws and treaties, and the actions and processes specific to politics (as opposed to those that are not exclusively political). Other types of political vocabulary, such as words used to express political values or to influence public opinion, are not the subject of this analysis. As institution vocabulary forms part of a specialist language, principles from terminology theory can be used to describe and categorise the terms. General strategies of term formation, as described for example by Sager et al. (1980) for English and Arntz et al. (2014) for German, can be adapted to group the Scottish Gaelic political terms from the Annual Reports into five broad categories:

- Borrowing: taking an entire term from another language, with or without assimilation
- Semantic extension: giving a new meaning to an existing word, thus adapting it for political terminology
- Derivation: term formation by prefixing or suffixing
- Compounding: combining several words to form a terminological unit
- Abbreviation: formation of an acronym or other short form, including as a synonym for another term

In addition to this categorisation, it makes sense to characterise the terms further by their relation to and use of other languages, especially English. Borrowing is a possible term formation strategy for any language, but for a minority language such as Scottish Gaelic it is closely linked to questions of language policy. Cooper (1989) points out that corpus planners are faced with a choice between coining new terms using the means of the target language, or borrowing from another language. However, as borrowed and native elements can be combined, the use of loanwords is considered in all types of terms here, rather than only in pure borrowings. Other foreign influences, such as loan translations, are also taken into account.

This analysis of the Scottish Parliament's terminology can be considered within the context of existing research on Scottish Gaelic corpus planning, terminology, and official registers. The Scottish Parliament's terminology was discussed by McLeod (2001) in his evaluation of the first parliamentary dictionary, *Faclair na Pàrlamaid*. While this dictionary and McLeod's evaluation cover a wider range of vocabulary than the political (institutional) terminology discussed in this paper, a comparison between the tendencies pointed out by McLeod (2001) and those found in the Annual Reports may be useful. As *Faclair na Pàrlamaid* was drafted in the early years of the Scottish Parliament, the analysis of the Annual Reports could point to possible developments in the terminology used by the Parliament. Since part of the Scottish Parliament's terminology has changed over the years, and since *Faclair na Pàrlamaid* and its successor *Faclair Rianachd Phoblaich* (2012) were never made available as a searchable database, they are currently not the primary terminology resources for translators working for the Scottish Parliament.[3] The subject of 'official Scottish Gaelic', not limited to politics or terminology, was also discussed by McLeod (2000), based on translated documents from the public sector. Beyond politics and state institutions, formal or specialised Scottish Gaelic registers have been discussed in other contexts, such as news scripts

[3] Alasdair MacCaluim (personal communication)

(Lamb 1999; Lamb 2008). McLeod (2004) discusses different aspects of linguistic purism in Scottish Gaelic, including the borrowing and creation of terms. Terminology issues are also addressed in research concerned with Scottish Gaelic corpus planning, e.g. the corpus planning study by Bell et al. (2014), with a survey on speakers' attitudes and preferences. While not concerned with the formation of individual terms, Bauer et al. (2009) provide a detailed review of terminology planning and resources for Scottish Gaelic, pointing out problems concerning co-ordination and dissemination. Studies like these indicate tendencies in Scottish Gaelic terminology use that can be compared to the results of this analysis of parliamentary reports. This will be discussed in the concluding section.

3. Types of term formation in the Annual Reports

There are some fairly clear tendencies regarding the frequency of term formation strategies used in the text corpus. While not the most common term category, loanwords from other languages make up a significant portion of the terminology. The vast majority of borrowed terms are assimilated to some degree. However, there is variation as to how far this assimilation goes. The more strongly assimilated terms are usually those that have been in use for a longer time, such as *lagh* 'law' and *pàrlamaid* 'parliament'. More recent loanwords are usually only adapted to conform to Scottish Gaelic spelling conventions without further morphological or phonological assimilation, e.g. *deamocrasaidh* 'democracy'. Nowadays, English is usually the source language for borrowings into Scottish Gaelic, even with words that are themselves loanwords in English (cf. Bell et al. 2014: 53). This is illustrated by international terms such as *deamocrasaidh*, *comataidh* 'committee', *pàrtaidh* 'party' and *poileataigs* 'politics'. The Scottish Gaelic forms mirror their English equivalents, which strongly suggests that these terms were borrowed through English. Despite the weaker assimilation of most loanwords, terms borrowed without any form of assimilation are very rare. Apart from acronyms and foreign proper names, there are only two examples in the text corpus: *Brexit* and the Irish loanword *reifreann* 'referendum', which seamlessly fits into Scottish Gaelic spelling conventions. Thus, while borrowing, especially from English, is a popular source of political terminology, there is apparently a reluctance to use more conspicuously foreign terms and to violate Scottish Gaelic orthography.

Semantic extension, i.e. giving an additional, specialised meaning to an existing Scottish Gaelic word, is a much less frequently used method of term formation in the examined reports. A few examples can be found, such as the following.

(1)

a. *coiteachadh*: a verbal noun meaning pushing or urging somebody, used in the reports to mean 'lobbying'
b. *gluasad*: a verbal noun for moving or motion, used to mean a motion in Parliament
c. *suidheachan*: a physical seat, used to mean a seat in Parliament

In the latter two cases, the semantic extension is the same as with the English terms *motion* and *seat*. It is thus likely that the terms are modelled after their English equivalents and are loan translations as well as examples of semantic extension. As a result, Scottish Gaelic and

English both have polysemic terms that are one-to-one equivalents in both contexts. It is worth noting that all cases of semantic extension in the text corpus are based on words still used in general language. There are no examples of the revival and adaptation of archaic words, in contrast to some political terms in Irish (see section 5).

The distinction between derivation and compounding in Scottish Gaelic is somewhat problematic. Words modified by 'prefixing' them with a noun or adjective are sometimes categorised as compounds rather than derivatives (Gillies 1993; Mark 2004). MacAulay (1992) does not distinguish between derivation and compounding at all and classifies combinations of a stem and affixes as compounds. A useful distinction for Scottish Gaelic is made by Csonka (2016), who differentiates compounds and derivatives based on the function of the added element rather than by whether or not it can be used independently. Derivation thus uses affixes that "[…] convey only logical meaning but no conceptual meaning […]" (42). When applying the latter definition, only a handful of terms from the Annual Reports fall into the category of derivation. However, some compound terms contain elements formed by derivation. For example, the adjectives *poilitigeach* 'political' and *pàrlamaideach* 'parliamentary', derived from the loanwords *poileataigs* and *pàrlamaid*, are used in compound terms such as *pàrtaidh poilitigeach* 'political party' and *taghadh pàrlamaideach* 'parliamentary election'. The following table shows some typical examples of term formation by derivation found in the text corpus: the word *reachd* 'law', 'statute' is used to form a cluster of terms for related concepts.

Table 2: *Examples of terms formed by derivation*

Scottish Gaelic term	Prefixes/suffixes used	English equivalent
reachdas	*-as* (abstract ending)	*legislation*
fo-reachdas	*fo-* 'under'; *-as*	*secondary legislation*
reachdadaireachd	*-adair* (agent noun); *-eachd* (abstract ending)	*legislature*

All of these terms are morphologically motivated and quite transparent due to the use of common affixes and the fact that designations derived from the same stem are used for related concepts. The same is true for the noun compound *bun-reachd* 'constitution', which combines *reachd* with *bun* 'base'. While *reachd* is a word of Gaelic origin, borrowed words are also used for derivation with Scottish Gaelic prefixes and suffixes.

(2)

a. *Prìomhaire* 'Prime Minister': long-established loanword *prìomh* 'prime', 'first' and suffix *-aire* (agent noun)

b. *fo-chomataidh* 'sub-committee': loanword *comataidh* and prefix *fo-* 'under'

These examples show that the loanwords are very much assimilated into Scottish Gaelic and are treated the same way as native words would be (e.g. the lenition after *fo-* in the second example).

While derivation is one of the rarer types of term formation in the text corpus, compounding is by far the most common. This is perhaps unsurprising given that many of the political concepts discussed in the reports are quite specific. Compounding thus seems particularly suitable to modify existing one-word designations and form terms for sub-types

of bills, committees etc. Compound terms are also the most structurally diverse term category discussed here. For instance, they can include elements formed by derivation, as in *pàrtaidh neo-riaghaltais* 'non-government party'. From a terminological point of view, any "[…] combination of two or more words into a new syntagmatic unit with a new meaning independent of the constituent parts" (Sager et al. 1980: 265) can be considered a compound. Consequently, some Scottish Gaelic genitive and prepositional phrases, such as the ones listed below, can be classed as compound terms. There are thus different subtypes of compound terms found in the Annual Reports. The first are simple noun + noun compounds:

(3)
a. *bun- reachd*
base law
'constitution'
b. *comataidh cuspair*
committee subject.gen
'subject committee'

The two terms above differ in that in example a), the first component modifies the second while in example b) the second component modifies the first, with the second component in the genitive case. This is generally a useful distinction when discussing Scottish Gaelic compounds (cf. Mark 2004: 646), and both types are quite common. However, the second type is more frequent in newer compounds while the first type is more typical for long-established compounds (cf. Csonka 2016: 84–86). This is reflected in the text corpus, where the second compound type is much more frequent than the first one. Example b) above is typical for the terminology in the text corpus in that the term appears to be modelled after its English equivalent. This will be discussed in more detail below.

Another compound term type is the combination of a noun with an adjective or participle, as illustrated by the following examples:

(4)
a. *ball roinneil*
member regional
'regional member'
b. *mòr- sheisean*
big session
'plenary session'
c. *cuspair dìonta*
subject protected
'protected subject matter'

These compound terms are also frequent in the text corpus. Again, the examples show the different structures possible in Scottish Gaelic: in term a) the adjective (derived from the term *roinn* for a parliamentary region) follows the noun, in example b) the adjective precedes it, similarly to a prefix. The former is much more frequent in the terminology from the Annual Reports. Term c) uses a participle rather than an adjective. Examples a) and c) once again illustrate the tendency of the compounds to be modelled after the corresponding English ones

(though still following Scottish Gaelic grammar rules). Example b) shows how English loanwords are integrated into Scottish Gaelic terms and treated as native elements would be, as with the lenition of *seisean* after the preceding adjective.

The last group of compound terms to be discussed here are genitive phrases with the definite article (as opposed to simple noun + noun compounds with the second component in the genitive) and prepositional phrases. These terms are the most complex compound terms found in the Annual Reports in that they usually consist of at least three or four elements (some of which are also used as terms in their own right) and they tend to combine different term formation techniques. Most often they are designations for specific organisational units, such as party names and proper names for political institutions:

(5)

a. *Pàrtaidh Nàiseanta na h-Alba*
party national art.f.gen Scotland.gen
'Scottish National Party'

b. *Coimisean air Ath-leasachadh Pàrlamaideach*
commission on reform parliamentary
'Commission on Parliamentary Reform'

As the other compound types discussed above, these terms also tend to reflect the formation of their English equivalents. There are some slight deviations, as with the name of the Scottish Conservative and Unionist Party, *Pàrtaidh Tòraidheach na h-Alba* (literally 'Tory Party of Scotland'). Compared to other term types in the text corpus, compound terms with genitive or prepositional constructions seem to be particularly prone to inconsistencies. This might be because they leave more opportunity for variation, e.g. by using different prepositions or different grammatical structures altogether. Inconsistencies in the terminology will be considered in detail in section 6.

The last type of term formation to be discussed here is abbreviation. Although abbreviations are often used as synonyms of other terms, it is worth looking at them as designations in their own right for the purpose of this analysis. While the Scottish Gaelic reports do contain a number of abbreviations (all acronyms), most of them are simply borrowed from English and therefore best categorised as loanwords. Interestingly, borrowed acronyms are often used even when a Scottish Gaelic full form also appears in the corpus. For instance, the Scottish National Party is referred to as *Pàrtaidh Nàiseanta na h-Alba*, but the English acronym *SNP* is used as the abbreviation, rather than one based on the Scottish Gaelic full form. The corpus contains only three acronyms based on Scottish Gaelic terms: *AE* for *Aonadh Eòrpach* 'EU', 'European Union', *BP* for *Ball-pàrlamaid* 'MP', 'Member of Parliament' and *BPA* for *Ball Pàrlamaid na h-Alba* 'MSP', 'Member of the Scottish Parliament'. Reasons for this can only be speculated on. In some cases (such as *SNP*) the English acronym was probably firmly established in everyday use and Scottish Gaelic media before the Parliament came to choose its terminology.

Having described the different term formation types in the Annual Reports, the remainder of this paper will take a closer look at some of the general tendencies that can be discerned.

4. English influence

Throughout all types of term formation, the terminology in the reports is strongly influenced by English. This influence takes different forms. Firstly, there is the prevalence of loanwords, which are almost all borrowed from or, in case of internationalisms, via English. While borrowed terms are almost always adapted in terms of orthography and inflected and lenited as native words, some contain morphemes that are not typical for Scottish Gaelic. For instance, the terms *deamocrasaidh* 'democracy' and *poileataigs* 'politics' have endings taken from English and not traditionally used to form abstract nouns in Scottish Gaelic. They also appear in loanwords outside political terminology, such as *matamataigs* 'mathematics' and *eaconomaidh* 'economy'. In the term *bhòt* 'vote', the initial consonant is lenited to recreate the /v/ sound of the English word, which would normally only occur in certain contexts for grammatical reasons. As borrowed terms make up a good part of the basic parliamentary terminology, they are used to form many other terms, such as compounds, to designate more specific concepts. For instance, *bile* 'bill' reappears in designations for different kinds of bills: *bile riaghaltais* 'government bill', *bile buill* 'member's' bill' and *bile prìobhaideach* 'private bill'. As a result, most terms used in the text corpus contain at least some element borrowed from or through English.

In addition to such direct borrowings, there is a very high number of loan translations in the terminology.[4] As mentioned above, some terms formed by semantic extension or derivation are apparently modelled after their English equivalents, taking the English term formation strategy and applying it using Scottish Gaelic and/or borrowed elements. However, loan translations are especially prevalent with the compound terms in the text corpus, which are almost always some form of calque, either from Scottish Gaelic components only or as blends with components borrowed from English. There could be several reasons for this. Firstly, there is the general dominance of English in Scotland and Scottish politics, and fact that the Annual Reports are translations from English, as are almost all official texts written in Scottish Gaelic. It is thus very likely that the English terminology would have a strong impact on its Scottish Gaelic counterpart. This side effect of asymmetrical translation situations has been pointed out with regard to Irish by Cronin (1995: 90), who maintains that "[…] translation is not a marginal but a central activity in the development of the minority language". Besides the general, and perhaps subconscious, English influence through translation, newer parliamentary terminology was deliberately created to provide equivalents for existing English terminology, as stated for instance in a foreword to *Faclair na Pàrlamaid* (McNeir 2001). In this situation, loan translation lends itself as a quick term formation strategy that would feel natural for many speakers. In addition to being a sign of English influence, loan translation is a way of achieving terminological transparency for readers who are more familiar with political terminology in English – which would include many native Scottish Gaelic speakers. A closer look at compound terms that are not calqued from English

[4] While it is not possible to retrace the exact process of term formation in every case and determine whether a term was actually created by loan translation, this is very likely when a term reflects the structure of its English equivalent, as English is the dominant language in Scottish politics.

shows that these are usually self-explanatory from their components, as illustrated by the following examples.

(6)

a. *sgìre- taghaidh*
district election.gen
'constituency'

b. *siostam còir- bhòtaidh*
system right voting.gen
'electoral franchise'

c. *tiomnadh cumhachd*
ceding power.gen
'devolution'

d. *bun- reachd*
base statute
'constitution'

For the term *mòr-chuid iomarcach* 'supermajority', which is not quite as transparent, the English translation is given in brackets in the Scottish Gaelic text. This approach is recommended in the Scottish Parliament's translation guidelines, for when the translator considers a term difficult to understand (Scottish Parliament 2019). All of this suggests that there is a high concern for the transparency of the Scottish Gaelic terminology, and this seems to affect the way terms are chosen and used in the reports.

As a more subtle sign of English influence, the Scottish Gaelic terminology makes the same terminological distinctions as English. For instance, as English uses different terms for the head of government of Scotland (*First Minister*) and the United Kingdom (*Prime Minister*), the same distinction is made between *Prìomh Mhinistear* and *Prìomhaire* respectively. Apparently, there is a preference for one-to-one terminological equivalents. While this means that the Scottish Gaelic terminology mirrors English on yet another level, it is practical from a terminological point of view as one-to-one equivalents can facilitate translation, terminology work and bilingualism within the same political system.

5. A brief comparison with Irish

For a further characterisation of the Scottish Gaelic terminology, it is worth comparing it to that of its 'sister language' Irish. As a closely related language that gained official status much earlier and has more institutionalised corpus planning, Irish has been suggested as a possible inspiration for Scottish Gaelic terminology planning. For instance, in his evaluation of *Faclair na Pàrlamaid*, McLeod (2001: 10) describes Irish as an "underused resource" that could have been a helpful model for the Scottish Parliament's terminology, but was not sufficiently utilised. A comparison of the terminology from the text corpus with that listed in the official Irish terminology database Téarma.ie suggests that this remains true today. Some similarities can be found between the Irish and Scottish Gaelic terminologies. A few terms, such as the following, are almost identical:

(7)
a. 'republic': ScG *poblachd*, Ir *poblacht* (derivation)
b. 'constitution': ScG *bun-reachd*, Ir *bunreacht* (compounding)
c. 'petition': ScG *athchuinge*, Ir *achainí* (semantic extension)

However, these similarities are not very frequent considering the close relationship between the two languages and the high number of cognates in less specialised registers. While they might be the result of Scottish Gaelic taking inspiration from Irish, they could also be coincidental. In some cases, both languages use loanwords from English, which leads to similar terms in Scottish Gaelic and Irish.

(8)
a. 'party': ScG *pàrtaidh*, Ir *páirtí*
b. 'commission': ScG *coimisean*, Ir *coimisiún*
c. 'vote': ScG *bhòt*, Ir *vótáil*

The text corpus only contains one term clearly borrowed from Irish: *reifreann* 'referendum'. This borrowing is suggested in the evaluation of *Faclair na Pàrlamaid* (McLeod 2001: 22) as a replacement for the term *referendum* listed in the dictionary. In this case, the suggestion to draw on Irish terminology as a resource was thus followed, but overall, Irish still seems to have very little influence on the Scottish Gaelic terminology used by the Parliament. A closer look at the differences between the two terminologies suggests that there are different preferences in Scottish Gaelic and Irish when it comes to certain term formation strategies. While Irish has revived some archaic words and adapted them as modern political terms, such as *Taoiseach* for the Irish head of government, there is no such example in the Annual Reports. The term used for the respective office in Scotland is *Prìomh Mhinistear*, a calque of the English *First Minister*. In several cases where Scottish Gaelic uses a loanword, the Irish equivalent is formed without borrowed elements, as in the following examples.

(9)
a. 'committee': ScG *comataidh*; Ir *coiste*
b. 'democracy': ScG *deamocrasaidh*; Ir *daonlathas*
c. 'minister': ScG *ministear*; Ir *aire*

In some instances, both languages use the same term formation type, but the Scottish Gaelic terms are still more strongly influenced by English than their Irish equivalents. With the terms for 'politics', for example, both languages use loanwords, but with different degrees of assimilation. The Irish term *polaitíocht* uses an Irish abstract ending while the Scottish Gaelic equivalent *poileataigs* uses an ending borrowed from English. The Scottish Gaelic and Irish terms for 'motion' are both formed by semantic extension, but different words were chosen to be adapted as political terms. Irish extends the meaning of *tairiscint* 'offer', 'bid'. Scottish Gaelic, on the other hand, uses *gluasad*, the verbal noun of *gluais* 'move', and thus a semantic loan from English. The Irish solution is arguably more semantically motivated than the Scottish Gaelic one and consequently transparent without relying on loan translation from English. A similar solution would have been possible in Scottish Gaelic, but a loan translation

was preferred. All these examples suggest that Scottish Gaelic political terminology is less purist than its Irish counterpart and more open to different forms of borrowing from English. It is important to bear in mind that many terms from the Annual Reports are specific to Scottish and/or UK politics and do not have exact Irish equivalents as the corresponding offices, institutions or other concepts do not exist in the Republic of Ireland. The number of terms from this corpus that can be compared to Irish is thus limited and the observations made here are based on a small number of examples. They do, however, correspond to previous research on Irish and Scottish Gaelic terminologies, e.g. the observation that Scottish Gaelic is generally not very purist in comparison to other European minority languages (McLeod 2004: 25). The usage (or lack of usage) of archaic terms as a terminology source has been cited as a major difference between Irish and Scottish Gaelic (McLeod 2004: 38). The differences between the Irish and Scottish political systems (resulting in a lack of exact equivalents for some concepts) also mean that in some cases Irish cannot serve as an inspiration for Scottish Gaelic. For concepts from the Scottish political system, there is of course always a corresponding English term, which might be an additional reason why English was frequently used as a resource while Irish was not.

6. Inconsistencies in the Scottish Gaelic terminology

The last notable aspect to be discussed here is the relatively high inconsistency of the Scottish Gaelic terminology. Inconsistencies can be found on different levels, and in some cases different versions of a term even appear within the same report. Often there is simply a variation in spelling, such as two Gaelicised spellings of *politics* (*poilitigs* and *poileataigs*). Other orthographic inconsistencies concern hyphenation and capitalisation. For instance, the Scottish Gaelic equivalent for *consul general* appears as *àrd-chonsail* and *Àrd Chonsal* (note the additional inconsistency in the spelling of *consa(i)l*), and *member's bill* is rendered as both *bile buill* and *Bile Buill*. Some inconsistencies go beyond orthographical variation. The Scottish Gaelic equivalent for *First Minister's Questions* appears in four versions differing in their use of prepositions, the genitive case, the definite article and, in the last case, capitalisation:

(10)

a.	*Ceistean*	*a'*	*Phrìomh Mhinisteir*
	questions	art.gen	First Minister.gen
b.	*Ceistean*	*don*	*Phrìomh Mhinistear*
c.	*Ceistean*	*dhan*	*Phrìomh Mhinistear*
	questions	to.art	First Minister
d.	*ceistean*	*Prìomh Mhinisteir*	
	questions	First Minister.gen	
	'First Minister's Questions'		

On rare occasions, the differences between alternative designations make it difficult to recognise them as referring to the same concept. This is arguably the case with the terms for 'legislative consent memorandum', which use the same structure but different components:

(11)

a. *meòrachan aonta reachdais*
memorandum consent legislation.gen

b. *gluasad cead reachdais*
motion permission legislation.gen
'legislative consent memorandum'

Some acronyms are also treated inconsistently. While the term *Ball Pàrlamaid na h-Alba* 'Member of the Scottish Parliament' is usually shortened to *BPA*, occasionally the English acronym *MSP* is used instead. *BPA* is sometimes treated like the full form with regard to lenition (e.g. *do BhPA*), in other instances it is left unchanged (*do BPA*). There are also two different plurals used in the corpus: *BPAan* and *BPA*. The frequent variation in the Scottish Gaelic translations contrasts with a very high level of consistency in the English reports, which may indicate that the translations are not as thoroughly edited for terminological consistency. A lack of comprehensive, standardised terminology resources for translators may also account for consistency issues in the Scottish Gaelic reports.

7. Conclusion and discussion

This analysis of three Scottish Parliament Annual Reports has shown some tendencies in the Parliament's Scottish Gaelic political terminology. Perhaps the most notable one is the strong influence from English, beginning with the borrowing of complete terms. As is generally the case with loanwords in Scottish Gaelic (cf. Gillies 1993: 222f.), more recent borrowings are less strongly assimilated than older ones. Still, there are only very few terms that are not assimilated at all and/or violate Scottish Gaelic spelling norms or morphology. Loanwords are frequently blended with Scottish Gaelic elements to form 'hybrid' terms. Even more frequent than English loanwords are loan translations, especially of compound terms. Purely Gaelic-based terms are not as numerous, but there are examples of creatively formed and morphologically or semantically motivated terms that are distinct from their English equivalents and reflect the differences between the two languages. The terminology bears little similarity to that listed in the official Irish terminology database, despite the close relationship between the languages and the fact that Irish has been put forward as a possible inspiration for Scottish Gaelic terminology development. A comparison of the Annual Report terms with their Irish equivalents suggests that the Scottish Gaelic terminology is more open to borrowing from English than its Irish counterpart. The lack of similarity to Irish terminology may be partly due to differences in the political systems of the two countries. As a consequence of the different term formation preferences and the Scottish orientation towards English terminology, the differences between Scottish Gaelic and Irish increase, at least within the specialist language relating to politics. Besides term formation preferences, the analysis has shown a number of inconsistencies in terminology use, suggesting that the Scottish Gaelic terminology is not as firmly established as the English.

These observations are mostly in line with those made in earlier research on Scottish Gaelic terminology. For instance, the reluctance to use Irish as a resource was already pointed out by McLeod (2001) in his evaluation of *Faclair na Pàrlamaid*. The Scottish Parliament's

preferences apparently have not changed much in this regard. A low interest in turning to Irish for inspiration, as well as opposition to reviving archaic words for modern usage, have been found to be common among Scottish Gaelic speakers in general (Bell et al. 2014: 189–200). The dominant term formation patterns in the Annual Reports seem to reflect these attitudes. The heavy use of English loanwords and loan translations has also often been commented on, for instance with regard to the Scottish Parliament (McLeod 2001), the language of news reports (Lamb 1999) as well as terms relating to modern life in general (e.g. McLeod 2004). Despite the frequent use of English loans, formal registers in Scottish Gaelic have been described as having some purist tendencies, in the sense that unassimilated borrowing is usually avoided (cf. Lamb 1999: 152; Lamb 2008: 143–145). This form of "moderate purism" (McLeod 2004: 37) is reflected in the Annual Reports, which are also examples of formal language. However, given the prevalence of English loanwords with minimal assimilation, loan translations, and terms containing borrowed elements, the terminology is not purist in the sense of an overall avoidance of English influence. On the contrary, it appears to be modelled closely on its English counterpart. While a comparison with previous studies on Scottish Gaelic terminology and formal registers in general provides some interesting insights, further research focussing specifically on political terminology used outside the Parliament (e.g. in the media or the education system) could allow for a deeper characterisation of Scottish Gaelic political terminology.

The findings of this terminology analysis can be placed within the general context of Scottish Gaelic corpus planning. As mentioned previously, the use of the language in political institutions such as the Scottish Parliament is a fairly recent phenomenon and there are very few structures in place to co-ordinate corpus and terminology planning. This lack of co-ordination has been pointed out as problematic (e. g. by Bauer et al. 2009 and Bell et al. 2014) and inconsistencies have been discussed, for instance, in the context of *Faclair na Pàrlamaid* (McLeod 2001) and other public-sector documents (McLeod 2000). The corpus planning situation probably accounts for some of the inconsistencies in the Annual Reports. It may also have influenced the choice of terms and the use of English loans. The Scottish Parliament's translation guidelines show a high concern for the transparency of terms, which is no surprise given the current corpus planning situation and the fact that many Scottish Gaelic speakers are not regularly in contact with political terminology in the language (e.g. through their work or the education system). As readers are more likely to be familiar with the English terminology, borrowing and loan translation from English might be the safest way of ensuring transparency. This could at least partly make up for the lack of co-ordinated terminology planning and dissemination and help make the Parliament's texts accessible to all speakers. However, the heavy use of loan translation and borrowing from English can be seen as a double-edged sword, as this strategy creates a Scottish Gaelic terminology dependent on English and mirroring English concepts and term formation patterns. At least within the bounds of political vocabulary, it thus reduces the difference between Scottish Gaelic and English – two languages belonging to different language groups and functioning quite differently on many levels. Due to its orientation towards English, the terminology does not make full use of the word formation options available in Scottish Gaelic and distinguishing it from English. This phenomenon, and the associated problem for minority languages, is not

limited to term formation but is intertwined with translation in general. Writing about translation into Irish, Cronin (1995: 90) observes:

> Translators in minority languages are […] placed in a classical double bind. If they translate allowing the full otherness of the dominant language to emerge in the translation, inviting rather than eliminating anglicisms from their Irish translations, then the language into which they translate becomes less and less recognisable as a separate linguistic entity capable of future development and becomes instead a pallid imitation of the source language in translatorese. On the other hand, if they resist interference and opt for target-oriented communicative translations that domesticate the foreign text, the danger is one of complacent stasis. Translation no longer functions as an agent of regeneration in the target language.

The fact that almost all official texts published in Scottish Gaelic are translations from English has been pointed out as problematic before (cf. McLeod 2000), and the prevalence of loan translation in Scottish Gaelic political terminology might be something to be aware of in this context. The self-explanatory Scottish Gaelic terms that are not modelled after their English equivalents show that there are other ways of achieving transparency, but this strategy requires much more time and effort than calquing from English, and this may not always be possible under the current circumstances.

References

Arntz, R., Picht, H., and K.-D. Schmitz. 2014. *Einführung in die Terminologiearbeit* (7th ed.). Georg Olms.

Bauer, M., Ó Maolalaigh, R., and R. Wherrett. 2009. Survey of Gaelic corpus technology. University of Glasgow. http://www.gaidhlig.scot/wp-content/uploads/2016/12/Bauer-et-al.-2009-Teicne%C3%B2las-corpais-CR09-03-Corpus-technology.pdf (Accessed 22/06/2021)

Bell, S., McConville, M., McLeod, W., and R. Ó Maolalaigh. 2014. Dlùth is inneach – final project report. Linguistic and institutional foundations for Gaelic corpus planning. Soillse. https://www.gaidhlig.scot/wp-content/uploads/2016/12/%C3%93-Maolalaigh-et-al.-2014-Planadh-corpais-CR12-03-Corpus-planning.pdf (Accessed 22/06/2021)

Bòrd na Gàidhlig 2007. The national plan for Gaelic 2007–2012. https://www.argyll-bute.gov.uk/moderngov/documents/s32076/National%20Plan%20for%20Gaelic.pdf (Accessed 29/01/2022)

Bòrd na Gàidhlig 2012. National Gaelic language plan 2012–2017. Growth and improvement. https://www.gaidhlig.scot/wp-content/uploads/2017/01/National-Gaelic-Langauge-Plan-2012-2017.pdf (Accessed 29/01/2022)

Bòrd na Gàidhlig 2018. National Gaelic Language Plan 2018–2023. https://www.gaidhlig.scot/wp-content/uploads/2018/03/BnG-NGLP-18-23-1.pdf (Accessed 29/01/2022)

Cooper, R. L. 1989. *Language planning and social change*. Cambridge University Press.

Cronin, M. 1995. Altered states: Translation and minority languages. *TTR 8*(1): 85–103.

Csonka, V. 2016. *Compounding, preposed adjectives and intensifiers in Scottish Gaelic* (Publication ID glathesis: 2016-8220) [PhD thesis, University of Glasgow]. University of Glasgow Enlighten: Theses.

Dunbar, R. 2010. Language Planning. In M. Watson, and M. Macleod (eds.), *The Edinburgh companion to the Gaelic language*, 146–171. Edinburgh University Press.

Gillies, W. 1993. Scottish Gaelic. In M. J. Ball, and J. Fife (eds.), *The Celtic languages,* 145–227. Routledge.

Gaelic Language (Scotland) Act 2005. 2005 asp 7. https://www.legislation.gov.uk/asp/2005/7/contents (accessed 02/09/2021)

Klein, J. 1989. Wortschatz, Wortkampf, Wortfelder in der Politik. In J. Klein (ed.), *Politische Semantik: Bedeutungsanalytische und sprachkritische Beiträge zur politischen Sprachverwendung,* 3–50. Westdeutscher Verlag.

Krochmann, L. 2019. Politikwortschatz im Schottisch-Gälischen – Verfahren der Wortschatzerweiterung am Beispiel des Annual Report (Aithisg Bhliadhnail) des schottischen Parlaments [Unpublished MA thesis]. Leipzig University.

Lamb, W. 1999. A diachronic account of Gaelic news-speak: The development and expansion of a register. *Scottish Gaelic Studies* XIX: 141–171.

Lamb, W. 2008. *Scottish Gaelic speech and writing: Register variation in an endangered language*. Cló Ollscoil na Banríona.

MacAulay, D. 1992. The Scottish Gaelic language. In D. MacAulay (ed.), *The Celtic Languages,* 137–248. Cambridge University Press.

Mark, C. 2004. *The Gaelic-English dictionary. Am faclair Gàidhlig-Beurla*. Routledge.

McLeod, W. 2000. Official Gaelic: problems in the translation of public documents. *Scottish Language* 19: 100–116.

McLeod, W. 2001. Faclair na Pàrlamaid: a critical evaluation. University of Edinburgh. http://www.poileasaidh.celtscot.ed.ac.uk/faclairrep.pdf (Accessed 22/06/2021)

McLeod, W. 2004. 'Feumaidh sinn a' Ghàidhlig a chumail *pure*': Problems of linguistic purism in Scottish Gaelic. In D. ó Riagáin, and T. Stolz (eds.), *Purism. Second helping. Papers from the conference on 'Purism in the age of globalisation', Bremen, September 2001,* 25–45. Brockmeyer.

McNeir, C. L. (ed.) 2001. *Faclair na Pàrlamaid. Dictionary of Terms*. The Scottish Parliament.

Sager, J. C., Dungworth, D., and P. F. McDonald. 1980. *English special languages: Principles and practice in science and technology*. Brandstetter.

Scottish Parliament 2019. Scottish Parliament translation guidelines. https://archive2021.parliament.scot/Gaelic/Scottish_Parliament_translation_guidelines_2019.pdf (Accessed 24/09/2021)

Scottish Parliament Annual Reports

Scottish Parliament. 2017. Annual Report / Aithisg Bhliadhnail 2016–2017. https://archive2021.parliament.scot/PublicInformationdocuments/Annual_Report_2016_to_2017.pdf (Accessed 24/09/2021)

Scottish Parliament. 2018. Annual Report / Aithisg Bhliadhnail 2017–2018. https://archive2021.parliament.scot/Annualreportsandaccounts/Annual_Report_2017-18.pdf (Accessed 24/09/2021)

Scottish Parliament. 2019. Annual Report / Aithisg Bhliadhnail 2018–2019. https://archive2021.parliament.scot/Annualreportsandaccounts/SPAnnualReport2018-19DPS.pdf (Accessed 24/09/2021).

LingBaW *Linguistics Beyond And Within 8 (2022), 114-132*

Language awareness as a prerequisite of literacy skills*

Anna Kucharská, Klára Špačková and Martina Šmejkalová
Charles University in Prague, Czech Republic

Abstract
Monitoring children's language ability levels makes it possible to predict their school achievement and identify potential difficulties. Identifying literacy difficulties in terms of specific learning disabilities is an important topic for elementary school pupils. Today, there is a general consensus based on research evidence that the diagnostic battery of literacy skills should always include a test of language abilities, through which the prerequisites for individual reading skills and at-risk readers can be identified.

A working group has created a diagnostic tool for language awareness that allows the morphological-syntactic (subtest A and B) and semantic (subtest C) language levels of Czech children of elementary school age to be assessed.

Data analysis examines how individual language skills correspond to reading performance – decoding and reading comprehension in different reading modalities (oral and silent reading) and how they correspond to other prerequisites for reading, such as listening comprehension. The paper introduces a new diagnostic tool using a quantitative methodology. It is based on data from the research project *Key Literacy Skills for Primary School Pupils* – a diagnostic battery, in which 881 elementary grade pupils (Grade 1 – Grade 5) participated. The statistical parameters of the new tool are presented and the relations between the *Language Awareness Test*, decoding skills, and oral and silent reading comprehension are analysed.

Keywords: literacy; language skills; phonology; morphosyntactic ability; prerequisites for reading; diagnostics; reliability

1. Introduction

In this paper, we address the issue of possible diagnosis of language ability in relation to developing literacy in young Czech school-age children. We draw on a standardization study (Kucharska et al. 2021) that focused on the development of diagnostic tools for reading skills, the prerequisites for reading, and the related influences on reading, with an emphasis on the area of comprehension.

* The project Klíčové gramotnostní dovednosti u žáků základních škol – testová diagnostická baterie [Key literacy skills for primary school pupils – diagnostic battery], n. TL01000365 (2018–2021) is co-financed by the Technology Agency of the Czech Republic within the Eta Programme.

Along with other language tests, the *Language Awareness Test* (LAT) had a significant role in this diagnostic battery. First, we describe its format and features and examine its psychometric properties. Our main goal is to investigate the relationship of children's morphosyntactic and reading abilities. But before that, we would like to mention a few facts that influence the development of reading literacy in Czech children.

1.1. Teaching reading in the Czech Republic

It is a well-known fact that the development of reading skills is not only dependent on traditional reading instruction at school. A child enters school with pre-literacy skills and the level of these skills will determine the success of the reading instruction. The development of pre-literacy skills is part of the pre-school curriculum. There are a number of skills that are developed, and we consider language to be essential. Monitoring (diagnosing) the level of language skills achieved and their potential development through intervention programmes is crucial to the positive development of reading skills.

The focus of the development of reading skills is the first five grades of elementary school. The child progresses through the stages (Helus 2012) of *beginning reading*, when the foundations of reading are being built, through the stage of *developing reading*, when the foundations of reading are already established, and then the child continues to improve in reading (speed, comprehension, reading strategies). The last stage is *practised reading* (Helus 2012). From middle school age onwards, the students should be able to learn through reading. They should also practise reading not only in relation to school but also as part of their leisure activities.

Formal reading instruction in the Czech Republic begins at age six. There are two main methods used in teaching reading – analytic-synthetic (syllabication) and genetic (spelling). Only 10% of children are taught to read using other methods concurrently (e.g. global reading, Sfumato sung reading, etc.).

The two methods used most extensively differ in several aspects. The analytic-synthetic method is based on learning the principle of syllable formation and subsequent syllable reading practice. This skill is gradually improved until the child reaches the stage of fluent reading. The genetic method does not work with syllables and works with the formation of words from phonemes. At first, children acquire the letters of the capital alphabet, form words, and experiment extensively with language itself (listening to stories, storytelling, vocabulary development, word games). The basis of the approach is that everything that happens in reading is linked to comprehension. The letters of the lowercase alphabet and written letters do not come into play until the second half of the first grade, when the child has already mastered the basics of reading.

Czech children tend to achieve disappointing results in international research (PISA, PIRLS). In 2001, they were, for example, rated below average. In a subsequent survey conducted in 2011, there was a slight improvement. These comparative studies show that mastering reading strategies tends to be most problematic, while work with texts and reading comprehension are weaker as well. There is therefore a great emphasis on the question of how to support the further development of literacy skills. One possibility is to enable guidance

workers in the education system who carry out diagnostic assessments of reading skills to have effective diagnostic tools to assess all the relevant factors that enter into the acquisition of reading in children with reading difficulties. Although we have a range of diagnostic tools in the area of reading at our disposal, these are more or less focused on reading aloud and pay more attention to decoding and less attention to reading comprehension. This is also the basis of the approach of a team of experts from the Department of Psychology at the Faculty of Education in Prague, who decided to create a comprehensive battery of reading skills that incorporates foreign research-based approaches and allows for the assessment of all relevant factors – reading prerequisites, decoding, and comprehension, as well as personal and environmental influences. We consider the comprehension of texts to be pivotal – we read in order to understand.

1.2. Supporting poor readers

The Czech Republic has a long tradition of supporting pupils with reading disabilities. Diagnostic tools to identify poor readers have existed since the 1930s, with the most significant development in the diagnosis of reading skills occurring in the 1980s (Matějček et al. 1987). Since the analytic-synthetic method of reading was used exclusively at that time, most attention was paid to the reading technique (decoding) in the construction of diagnostic reading tests. Much less attention was paid to reading comprehension and other reading-related skills. Although children with specific learning disabilities received appropriate support within the emerging model of counselling support, the visual model of dyslexia (a reading problem as a consequence of a visuomotor handicap) was favoured in the interpretation of the nature of specific learning disabilities, which subsequently influenced interventions for poor readers.

Unfortunately, this approach still more or less persists to this day, with a relatively large group of practitioners involved in diagnostic and remedial reading instruction still considering reading to be a visuomotor skill. Although new diagnostic methods of reading are emerging that gradually take into account the linguistic background of specific learning disabilities, they make little use of the assessment of reading comprehension. When they do, the preferred approach is the "retelling" of the story after reading it, which can be subjective. Poor performance in retelling may be due to reasons other than comprehension problems (personal set-up, fear, stage fright, level of expressive skills, etc.).

1.3. Diagnostic battery to assess reading skills in younger school-age children

As a reaction to the limits of the current Czech assessment measures, the new test battery *PorTex* was designed. The diagnostic battery was prepared for younger school-age children in grades 1 to 5, a period that is crucial in terms of developing reading skills. The advantage of the battery is its comprehensiveness. The individual tests are grouped into thematic clusters according to common characteristics and an overview of the tests is included in the Appendix.

Unlike any other Czech test battery, *PorTex* focuses extensively on text comprehension. Various comprehension task formats have been included to highlight the developmental nature of comprehension being built (see Figure 1):

- Listening is an implicit skill developed even prior to starting school. We must first understand the spoken word in order to understand a read text. Children with better language skills are likely to have better listening skills.
- At the beginning of the development of reading, reading comprehension is strongly linked to decoding. Children with decoding problems also have poorer comprehension. As literacy skills develop, even children with poorer decoding skills may attempt to understand texts if they identify with the function of reading (reading gives us information and allows us to live through an experience).
- Silent reading comprehension is built in conjunction with reading aloud; it is the beginning of the development of reading strategies. It can allow for better comprehension when a child is stressed by reading aloud.

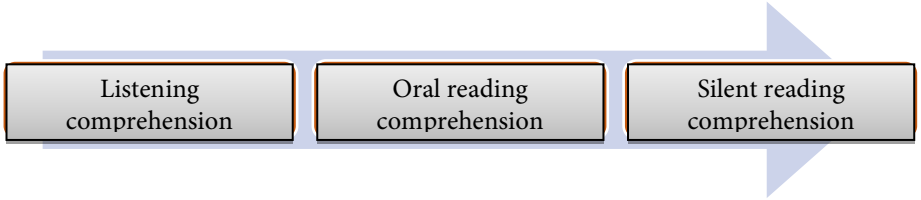


Figure 1: *Development of Comprehension (Kucharská et al. 2021)*

In developing the *PorTex* diagnostic battery, we drew on the theoretical framework of the Simple View of Reading (Gough and Tunmer 1986), which views reading comprehension as the result of two related but somewhat independent components: decoding skills (the ability to read printed words) and linguistic comprehension (the ability to understand oral language). The model thus draws attention to the fact that a good level of decoding is not sufficient for a good understanding of a text, just as linguistic comprehension alone is not sufficient. Both of these components, which are projected into reading performance, can be assessed using diagnostic tools – decoding tests and listening comprehension tests (Hoover and Gough 1990).

The test battery is based on the principle that reading cannot be approached as a monolithic process. The assessment must make a distinction between decoding and comprehension of the reading text. When measuring the level of these sub-skills of reading, however, it is also necessary to map core skills – that is, the skills that are crucial to their development. According to the current state of knowledge, the following core prerequisites constitute language skills.

It is now widely known that phonological awareness across alphabetic languages is closely linked to the development of decoding skills (e.g. Caravolas et al. 2012). Measuring phonological awareness has thus become a common part of the diagnostic process of reading skills. However, this is not the case for language awareness at higher levels of language. The influence of morphosyntax and semantics has been documented primarily in relation to reading comprehension (Tylor and Nagy 1900; Foorman, Petcher, and Bishop 2012; Snowling and Hulme 2012) but test batteries of reading skills for first-graders do not offer measurement scales for these language skills, at least not in the Czech Republic.

If we look at morphological awareness and its role in the development of reading skills in more detail, the longitudinal study of Kirby et al. (2012) came up with interesting findings. In a research sample of 103 younger elementary school pupils, morphological awareness provided an explanation for the variance in reading comprehension even after controlling for the effects of word reading. Furthermore, it also had an effect on word reading accuracy and efficiency above and beyond intelligence and phonological awareness. However, as is the case in several other research studies (e.g. Deacon and Kirby 2004), this relationship did not occur with first-graders. Thus, the role of morphological awareness appears to change over time.

1.4. Aim of the study

In this paper, we set out to introduce one of the language tests in our test battery, namely the *Language Awareness Test (LAT),* which was designed to assess a child's level of morphosyntactic and lexical-semantic abilities. Other language tests include the *Phonemic Awareness Test,* focusing on phonology, and the *Pseudoword Repetition Test*, but these are not presented in this study.

We will describe the test itself and its developmental characteristics in terms of performance and evaluate its psychometric properties. We will examine whether children being taught with the two most commonly used reading methods differ in their performance on the LAT, and whether performance differs by gender. Attention will also be paid to correlation analysis between the LAT and reading tests. Finally, further intentions with this method will be presented.

2. Methodology

2.1. Standardization study

The diagnostic battery was standardized through a project of the Technology Agency of the Czech Republic in 2018–2021 (Kucharská et al. 2021). The main goals of the project were to broaden the view of the diagnosis of reading skills, prepare instruments, and standardize them for use in schools and counselling practice.

The data collection was conducted in two phases in 2018 and 2019, in the first half of the school year (T1) and in the second half of the school year (T2). The sample of children covered the entire Czech Republic. The schools where data collection took place were divided according to socio-economic factors (unemployment rate in the region) into ten clusters and the sizes of the schools were also taken into account, in addition to the socio-economic level in the area. The total number of children whose performance entered into the development of the norms is shown in Table 1.

Table 1: *Participants*

GRADE	T1 (M)	T2 (M)	TOTAL (M)
1	121 (49%)	129 (52%)	250 (50%)
2	202 (48%)	186 (47%)	388 (48%)
3	213 (40%)	186 (42%)	399 (41%)
4	197 (40%)	190 (40%)	387 (40%)
5	148 (47%)	160 (46%)	308 (47%)
TOTAL	881 (44%)	851 (45%)	1732 (45%)

Note: These are the total numbers of children included in the study. The numbers of children whose performance figures in the analyses may vary slightly. This is because a given test was not always administered to a child, e.g. because of the child being absent at the time of the test. Such a child was not excluded from the study; all available tests were used.

2.2. Methods used

2.2.1. Language Awareness Test (LAT)

The test was developed after a pilot investigation in 2014–2015 (Kucharská et al. 2015) with the aim of assessing morphosyntactic and lexical-semantic abilities in younger school-age children (first to fifth grades). For standardization, which was ensured through the TACR project, it underwent a partial transformation; non-differentiating items were removed. This resulted in a reduction of the total number of items (from 60 to 48). The items to be included in the final version of the test were selected so that adequate percentage amounts were covered while allowing for the option to increase the level of difficulty, so that the test could work both for younger and older elementary school pupils in grades 1 to 5.

The test currently contains three subtests: *Morphology, Word Formation,* and *Sentence Comprehension.* Each section includes four subtests with four items, for a total of 48 points. The test is administered orally; each item includes a practice session, in which the subject becomes familiar with the type of task and the method of response. For most questions, the children answer in their own words; in only four tasks do the children choose from a set of possible answers (B1, C1, C3, C4).

The subtests of individual parts include the following tasks:

A. Morphology

- **A1 Grammatical Number.** Our task was to demonstrate awareness of the opposite of singularity and plurality (on material more difficult than the one banana x two bananas model). The words used are singularia tantum, i.e. words that morphologically have only a singular form but denote multiple items. Thus, these are cases where singular nouns denote more than one entity. The aim is to find out whether the child's morphological awareness of the grammatical feature "singular number" overrides the actual number of entities denoted.
- **A2 Case.** The tasks on the conscious use of the case grammatical category are a standard part of language awareness tests and verify the subject's orientation in the relatively complex situation of Czech as an inflected language which is characterized by a plethora

of seven different cases (in singular and plural forms as well as feminine, masculine, and neuter). In our test, we focused on the declension of pronouns (in Czech, other so-called flexible verbal types, such as adjectives, numerals and pronouns, are also declined). The ability to form the correct case is mainly related to the clear expression of the syntactic function of the noun in question.

- **A3 Voice.** In creating the task, we were inspired by examples that have recently been introduced, e.g. in the *Encyclopedic Dictionary of the Czech Language* (Karlík 2017): "[Voice (genus verbi) is] a grammatical category whose typical representative is the opposition represented by the structures (a) The teacher praises the pupil and (b) The pupil is praised by the teacher." The main purpose of the task was to find out whether the children were aware of the different possibilities "what semantic participant in the event is expressed in the syntactically prominent position of the subject" (Kucharská and Šmejkalová 2017) while fundamentally changing the "external" (from the child's point of view) form of the predicate while preserving the basic meaning of the sentence. Simplified and modelled, this is what is described in traditional Czech linguistics as a situation (ibid.) in which "the originator of the action in the syntactic structure of the sentence is represented either by its subject" (the active or active voice; see structure (a)) or is "expressed in a different syntactic position" (the passive or the passive voice; see structure (b)). The basic meaning of the sentence is preserved.
- **A4 Verbal aspect.** This task tested whether the child could distinguish the fact that "the Czech verb exists in two (up to three) forms that have the same lexical meaning but differ from each other in their relation to the completion of the action" (Grepl et al. 2008: 318). Thus, we respect the generally prevailing notion according to which imperfective verbs "express a loose relation to the completion of the action," while perfective verbs express "the limited duration of the action" in the sense of "completeness, wholeness of the action," where they denote the fact that "the goal has been achieved and there was no point in continuing" (ibid.: 318–319). This category can be supplemented by one of the more recent definitions of verbal aspect: "In general linguistics, the international term aspect denotes a more or less strongly grammaticalized category of the verb that expresses the non-temporal time dimension of the action/state expressed by the verbal lexeme; hence, we speak of verbal aspect" (Nübler et al. 2017).

B. Word Formation

- **B1 Noun formation.** This is an uncomplicated test task to determine whether the child has consciously associated less common suffixes with the word formation of the relevant noun types. The principle of the test task was developed by Hájková (see Hájková et al. 2013). In the task, typical suffixes for deriving nouns were investigated. As an example, the suffix which creates place names (denominatives and deverbatives) such as *cukrárna*/pastry shop, etc. (Dokulil et al. 1986: 235, 284). Since the word formation of nouns is relatively stable in Czech, children should gradually acquire a repertoire of word formation devices automatically.

- **B2 Adjective formation.** Again, this is an uncomplicated test task to see if the child has consolidated the word formation characteristics of (and differences between) meaningfully different types of adjectives. The task is to form an adjective based on information from the previous sentence. As the relevant literature states, "the basic features of adjective derivation arise from the naming character of adjectives" (Dokulil et al. 1986: 316). Here, we built on the earlier research conducted by Šmejkalová (2012; the experimenter was Babušová) that third-grade children already spontaneously label adjectives as "characteristics," which intuitively corresponds to the grammatical and functional-semantic characteristics of adjectives as postulated in the expert literature. We therefore expected relatively developed prerequisites for solving this task and the research results confirmed this expectation
- **B3 Comparative and superlative adjective formation.** It is assumed that children entering formal schooling have a good grasp on grading adjectives, and therefore the test task was made more difficult by the element of suppletion. The solution of the task is strongly dependent on the level of the child's vocabulary. In addition to the regular grading of adjectives, we made the grading of adjectives more difficult by the manifestation of suppletion ((dlouhé – delší – nejdelší/long-longer-longest, dobré – lepší – nejlepší/good-better-best) in order to increase the difficulty of the test. Suppletion can be defined as "a type of allomorphy, in which the relationship between forms of one morpheme is not derivable by phonological (or morphophonological) rules. These irregular forms are referred to as suppletive allomorphs" (Ziková 2017).
- **B4 Adverb formation.** The task involved the formation of adverbs. This is an easier task, especially because "adverbs as a whole are characterized by less word formation specificity and a shortage of derivational devices" (Dokulil et al. 1986: 431) and their formation is largely paradigmatic. That is why we included more demanding adverbs from compound adjectives, which were qualitative, whose linking vocalic is -o-, and which are converted by the formant e/ě (ibid.: 444).

C. Sentence Comprehension

- **C1 Word order**. This is a more difficult syntactic task and a phenomenon in which even adult speakers make mistakes, the so-called false syntagma. It is an inappropriate word order resulting from the interpretation of "a pair of sentence elements as a syntagma when this interpretation is not intended." The aim was to verify whether, or at what age, children are able to distinguish this subtle difference in meaning and apply a more appropriate word order.
- **C2 Topic-focus articulation (sorting utterances).** Although the syntactic literature treats issues of topic-focus articulation differently, we based this task on the theory presented by Grepl and Karlik (1998: 495). Here, the topic-focus articulation is seen as "the division of the content of the utterance into an initial part and a core part," with the initial part being understood as "the part of the utterance that usually expresses what is known in a given situation" and is the lesser component in terms of communicative weight. The core is then understood by the authors as "that part of the utterance which expresses the facts of greatest communicative weight" (ibid.: 495). We observed whether

children identify the core of the utterance as its most communicatively serious part, referring to different communicative contexts and illocutionary consequences; we based this on the model presented in the cited literature (*Peter sent the principal **a letter**. – The letter was sent to the principal by **Peter**. – Peter sent a letter to the **principal**.* – etc., ibid.: 496).

- **C3 Sentence structure – verb valency.** Verb valency is a concept whose understanding has evolved and continues to evolve, as has the development and understanding of syntactic valency theory. We build on the obvious fact that sentences "such as *the mother met, the director dictated, the car was heading, etc.* are nonsensical and therefore non-grammatical" (Hirschová 2012: 8). Our aim was to determine whether children perceive this ungrammaticality. We therefore relied on the notion of the so-called academic Czech Grammar in Use textbook, which indicates – to put it simply – that valency is the ability of a dominant element, usually a verbal predicate, to require a certain number of its complements, to which it also assigns a form (Daneš, Hlavsa, Grepl, et al. 1987: 18). We observed whether the children have a well-established awareness of the form of the syntactic core of the sentence (Hirschová, ibid.).
- **C4 Communicative function of utterances.** The last item tested was the children's ability to identify the likely communicative functions of an utterance, i.e. whether they could identify the current communicative function: "why, with what aim (intention, purpose) it is said" (Grepl and Karlik 1998: 333), in addition to the propositional content of "what is said". We induced the communicative context and observed whether the children distinguished the probable CF. We made the task more difficult by contrasting the probable CF with the learners' preconceptions about the sentences according to the speaker's attitude in some cases ("Break a leg!"); for example, we deliberately formalized the utterance with the CF of warning as a so-called command sentence, and we also involved tropes such as irony.

2.2.2. Comprehension tests

The comprehension tests were constructed on the basis of the same principles in three modes: listening comprehension, oral reading comprehension, and silent reading comprehension. We respected the developmental needs of the children in terms of the content of the texts, as well as in relation to the reading levels they had achieved. We therefore worked with three versions of the tests: version A for Grade 1 pupils, version B for children from Grades 2 and 3, and version C for children from Grades 4 and 5.

One third of the questions in the tests was focused on what is called literal comprehension (explicit comprehension), in which the extent to which the child has understood the information from the text (read, heard) and can recall it by answering questions (open-ended and multiple-choice) is monitored. The second third of the questions was directed towards inference (implicit comprehension). In order to answer correctly, it was no longer sufficient to understand the text, but it was necessary for the child to be able to draw certain conclusions from the information – to be able to "read between the lines." We assume that other cognitive abilities and skills are reflected in these differently focused areas of understanding. The last

third of the questions focused on the interpretation of texts. Here again, the child's other skills and personality traits, as well as their background knowledge, were reflected.

The questions were posed after listening to or reading a text. They were mostly open-ended; the child answered the questions about the text in their own words. For silent reading comprehension, questions were created in which the child chose from a list of possible answers.

In total, a maximum of 30 points could be obtained in each test, while the partial scores for explicit and implicit comprehension and interpretation of the text amount to a maximum of ten points in each test.

Comprehension tests are based on linguistic assumptions. We therefore supplemented all the tests with an orientation vocabulary test (*Control vocabulary test*), which consisted of terms that occur in the text. This was to check that low performance on the test was not due to poor understanding of the words. For this reason, words that were not very common or rather more challenging for a given age level were selected.

2.2.3. Decoding

Oral reading comprehension tests allow for the assessment of decoding in addition to comprehension. The child is asked to read aloud the texts we work with. Several indicators are monitored during the reading process to inform us about the level of reading attained. In this text, we use three of them:

1. the number of words read correctly in the first minute,
2. the total time taken to read the stimulus text, and
3. an accuracy score (not in terms of the number of words read incorrectly, but in terms of their percentage in relation to the total number of words read)

3. Results

3.1. Test performance by children in grades 1 to 5

Table 2 and Figure 2 show the performance characteristics of the *Language Awareness Test* (LAT) from grades 1 to 5 in elementary school.

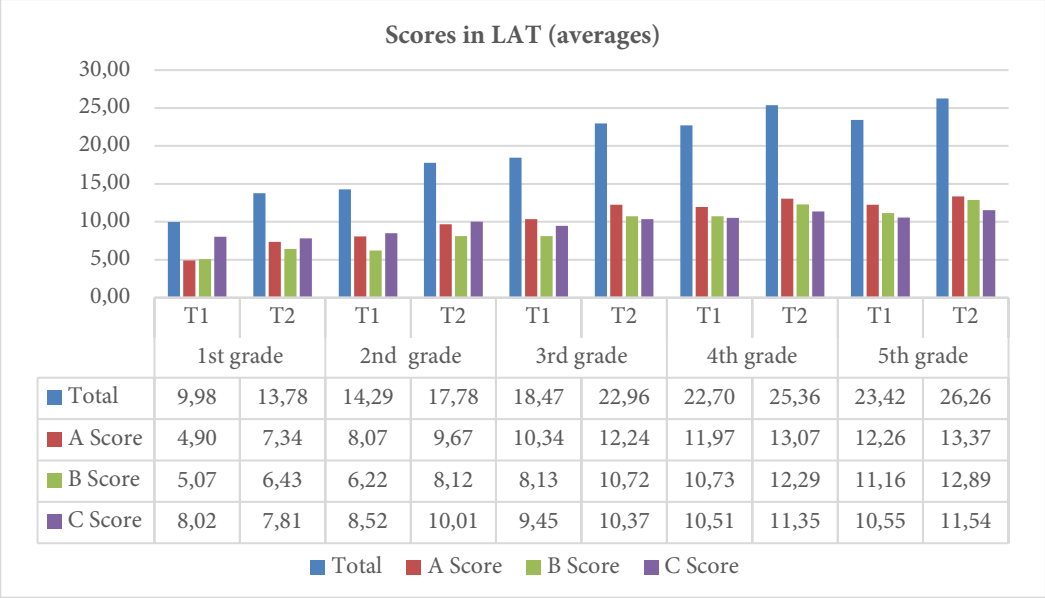
The item analysis (Table 3) shows the average scores achieved in each subtest and partial tasks (the maximum score is one point). In some items, especially in parts A and B, it was possible to set the order with a developmental emphasis – i.e. the easier items appear at the beginning of the subtest, the harder items at the end of the subtest. The other interesting trend in the results achieved is that, with some exceptions, success rates with the testing stages and in the upper grades are higher. This is particularly true for Grades 1–3, with the performance of older children no longer showing large differences and fluctuating across testing stages and grades. Subtest C is the weakest in terms of both tendencies.

Table 2: *Item analysis – means*

	Grade 1		**Grade 2**		**Grade 3**		**Grade 4**		**Grade 5**	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
N	119	102	198	170	212	164	196	181	148	149
A1_1	0.34	0.52	0.58	0.71	0.78	0.87	0.87	0.97	0.90	0.93
A1_2	0.29	0.30	0.47	0.55	0.69	0.84	0.89	0.93	0.86	0.95
A1_3	0.50	0.70	0.68	0.78	0.84	0.91	0.88	0.94	0.84	0.95
A1_4	0.24	0.26	0.40	0.53	0.59	0.72	0.76	0.83	0.76	0.87
A2_1	0.57	0.75	0.88	0.85	0.86	0.88	0.86	0.86	0.86	0.87
A2_2	0.39	0.62	0.71	0.85	0.71	0.89	0.78	0.93	0.76	0.85
A2_3	0.29	0.36	0.49	0.62	0.63	0.72	0.70	0.79	0.76	0.81
A2_4	0.09	0.22	0.34	0.34	0.50	0.57	0.70	0.83	0.74	0.78
A3_1	0.13	0.23	0.30	0.46	0.56	0.79	0.80	0.80	0.82	0.83
A3_2	0.26	0.57	0.46	0.50	0.63	0.80	0.72	0.76	0.72	0.74
A3_3	0.33	0.43	0.54	0.56	0.61	0.79	0.70	0.77	0.74	0.79
A3_4	0.07	0.10	0.13	0.31	0.26	0.49	0.47	0.61	0.50	0.76
A4_1	0.50	0.53	0.58	0.78	0.78	0.85	0.85	0.88	0.81	0.89
A4_2	0.41	0.55	0.67	0.72	0.78	0.79	0.76	0.80	0.75	0.87
A4_3	0.34	0.50	0.56	0.69	0.70	0.80	0.74	0.79	0.82	0.84
A4_4	0.15	0.21	0.28	0.39	0.43	0.55	0.46	0.56	0.59	0.68
B1_1	0.59	0.58	0.64	0.75	0.71	0.85	0.82	0.91	0.84	0.95
B1_2	0.64	0.75	0.70	0.78	0.82	0.92	0.93	0.95	0.95	0.95
B1_3	0.35	0.40	0.33	0.46	0.56	0.74	0.69	0.80	0.74	0.77
B1_4	0.48	0.52	0.55	0.53	0.57	0.72	0.62	0.70	0.68	0.77
B2_1	0.45	0.51	0.49	0.62	0.63	0.74	0.80	0.88	0.72	0.87
B2_2	0.09	0.19	0.15	0.26	0.25	0.51	0.61	0.67	0.61	0.82
B2_3	0.18	0.27	0.24	0.44	0.50	0.71	0.80	0.85	0.80	0.92
B2_4	0.13	0.28	0.21	0.35	0.33	0.42	0.51	0.49	0.49	0.63
B3_1	0.20	0.25	0.26	0.42	0.37	0.61	0.57	0.73	0.68	0.84
B3_2	0.18	0.24	0.22	0.36	0.29	0.59	0.54	0.75	0.64	0.86
B3_3	0.18	0.26	0.24	0.37	0.39	0.57	0.49	0.75	0.61	0.76
B3_4	0.10	0.12	0.16	0.24	0.18	0.30	0.27	0.41	0.37	0.62
B4_1	0.34	0.51	0.60	0.79	0.81	0.91	0.91	0.94	0.82	0.97
B4_2	0.47	0.57	0.60	0.77	0.79	0.93	0.92	0.96	0.90	0.96
B4_3	0.62	0.69	0.74	0.86	0.76	0.90	0.80	0.87	0.80	0.84
B4_4	0.06	0.09	0.10	0.18	0.22	0.35	0.45	0.57	0.52	0.57
C1_1	0.43	0.39	0.53	0.45	0.59	0.55	0.68	0.57	0.65	0.62
C1_2	0.76	0.74	0.63	0.68	0.54	0.67	0.61	0.71	0.67	0.71
C1_3	0.53	0.48	0.65	0.52	0.66	0.70	0.73	0.73	0.79	0.81
C1_4	0.65	0.55	0.45	0.50	0.41	0.41	0.33	0.40	0.36	0.50
C2_1	0.32	0.43	0.55	0.58	0.60	0.69	0.71	0.80	0.68	0.81
C2_2	0.42	0.61	0.50	0.61	0.66	0.77	0.76	0.86	0.73	0.87
C2_3	0.13	0.11	0.17	0.22	0.27	0.49	0.46	0.52	0.41	0.56
C2_4	0.15	0.19	0.12	0.19	0.28	0.39	0.44	0.56	0.39	0.56
C3_1	0.36	0.37	0.46	0.45	0.47	0.47	0.52	0.54	0.54	0.58
C3_2	0.90	0.85	0.93	0.98	0.98	0.98	0.98	1.00	0.99	0.99
C3_3	0.88	0.92	0.97	0.97	0.98	0.98	0.99	0.99	0.97	0.99
C3_4	0.22	0.14	0.13	0.12	0.11	0.16	0.17	0.20	0.20	0.27

C4_1	0.71	0.78	0.79	0.86	0.80	0.86	0.84	0.92	0.84	0.87
C4_2	0.59	0.50	0.45	0.51	0.62	0.74	0.63	0.81	0.72	0.81
C4_3	0.46	0.53	0.55	0.66	0.69	0.75	0.76	0.81	0.72	0.80
C4_4	0.52	0.47	0.64	0.69	0.81	0.89	0.90	0.90	0.89	0.93

Figure 2 describes the average values in each LAT subtest. It also shows the developmental character of performance – the trend of gradually increasing performance in the individual subtests (A, B, C) of the test.



	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
	1st grade		2nd grade		3rd grade		4th grade		5th grade	
Total	9,98	13,78	14,29	17,78	18,47	22,96	22,70	25,36	23,42	26,26
A Score	4,90	7,34	8,07	9,67	10,34	12,24	11,97	13,07	12,26	13,37
B Score	5,07	6,43	6,22	8,12	8,13	10,72	10,73	12,29	11,16	12,89
C Score	8,02	7,81	8,52	10,01	9,45	10,37	10,51	11,35	10,55	11,54

Figure 2: *Average values in the total LAT score, in the scores of subtests A, B, and C*

3.2. Reliability

When designing a diagnostic tool, it is necessary to check its reliability. We assessed it through the internal consistency coefficient (Cronbach's alpha). We were also interested in the test-retest reliability using the correlation coefficient between the results in T1 and T2.

Table 3 shows the reliability (Cronbach's alpha) of each subtest for testing stages and grades. The results show that the values are sufficient for subtests A and B, while insufficient reliability was identified for subtest C (Kline's rule states that it should be at least 0.7).

The reasons for the low reliability (i.e. low correlation of items with the overall test score) are apparent in several factors. Table 3 shows that some items have high performance across all grades (e.g. C2.2, C2.3) but it is also clear that there is no ascending trend between grades and stages of testing as in subtests A and B. This could be due to either poorly chosen tasks (subtests) to represent sentence comprehension or poorly chosen tasks within them, or it could be in relation to the way the questions were posed. Unlike the previous tasks, here the children were largely asked to choose from a list of answers, and thus they could guess rather than actually answer. Intonation could also have played a part here – in the way the sentence was pronounced – and it may be that it was the intonation that in fact cued the child to the correct answer without lexical-semantic links. We have therefore decided to exclude this subtest from the standards we are preparing. It is necessary for the subtest to undergo further

revision. However, we are not abandoning the domain as such – sentence comprehension is an important aspect of language awareness that we would like to develop further in the future. For these reasons, we will not address Subtest C below.

Subtests A and B have sufficient reliability and on aggregate, when considered across grades and stages of testing and working with only these two subtests, it reaches a value of 0.9 (removing one subtest increases the internal consistency of the test).

The test-retest reliability assessed by means of correlation coefficients (Pearson's *r*) between stages T1 and T2 reaches a value of 0.85, which is a satisfactory result.

Table 3: *LAT Reliability (Cronbach's alpha)*

	Grade 1		**Grade 2**		**Grade 3**		**Grade 4**		**Grade 5**	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Total	0.72	0.75	0.78	0.83	0.80	0.81	0.77	0.80	0.83	0.81
Subtest A	0.70	0.68	0.71	0.72	0.71	0.71	0.62	0.63	0.75	0.70
Subtest B	0.66	0.71	0.70	0.74	0.73	0.70	0.67	0.67	0.74	0.76
Subtest C	0.31	0.11	0.21	0.42	0.34	0.41	0.40	0.45	0.43	0.33
Reliability (after corrections)										
Total	0.78	0.82	0.8	0.84	0.81	0.81	0.75	0.76	0.83	0.83
Subtest A	0.70	0.72	0.71	0.72	0.71	0.70	0.62	0.62	0.75	0.74
Subtest B	0.60	0.71	0.70	0.69	0.75	0.73	0.69	0.67	0.74	0.75

3.3. Differences in performance in the LAT between reading instruction methods and gender

When constructing diagnostic tools, it is also necessary to monitor whether there are differences in performance between children in terms of belonging to certain groups. It would seem that there may be differences between various methods used for teaching reading, and one cannot overlook gender differences.

The LAT points to implicit language abilities that develop without intentional development in preschool. At school age, however, it is not through teaching alone that schools can have an impact. It is also about how teaching encourages language development. The genetic method of reading builds more on the child's language competence and develops it in reading instruction. This was also in fact reflected in the results obtained (see Table 4), where in the overall summary, children taught with the genetic method reached better results, with a medium effect size (−0.41). Looking at the results in more detail, we see (outside the table) that this applies to all three scores – the overall score, the score for subtest A, and the score for subtest B. If we look at the differences across the grades, it is especially apparent in first grade at T2, which may indicate that the method may already have this effect after six months of application. In the other two grades, when the foundations of reading are established and the child is at the stage of developing further reading skills, we found only minor and nonsignificant differences between the methods. And subsequently, the differences between the methods are already identified in Grades 4 and 5. At this stage, the child is already practising reading and it is possible that children taught with the genetic method are more successful in this respect and benefit more from reading and the teaching method.

The findings have also contributed to the way in which the standards are handled. We have a procedure in place that will allow for some correction of the results precisely with regard to the method used for teaching reading.

The results of the international comparative studies PIRLS (2011) and PISA (2015) show significant differences in reading test scores between girls and boys; girls are statistically significantly better readers than boys. Authors in Anglo-Saxon countries, such as Robinson and Lubienska (2011) in the USA, have also reached similar results using data from a longitudinal study of reading in children from grade 0 (K – Kindergarten) to grade 8. According to them, girls' improved performance is evident at the beginning of schooling and persists throughout compulsory schooling. The same findings have been echoed in the UK, where the focus has been on identifying early reading strategies and developing them specifically in boys to reduce the gap between boys and girls (DfES 2009; Price-Mohr and Price 2017). We were therefore also interested to see whether the differences would be reflected in the LAT (see Table 5).

As is apparent in Table 4, there were no gender differences in the level of linguistic ability in the area of morphosyntactic processing. The boys and girls in our test battery show similar aptitudes for the development of reading competence.

Table 4: *Differences between Reading Teaching Methods and Gender in the LAT*

	method	**N**	**IRT score**		***t*-value**	**degrees of freedom (*df*)**	***p*-value**	**Cohen's *d***
			arithmetic mean	**standard deviation**				
Reading Teaching Method	AS	1276	−0.07	0.93	−7.27	1691	<0.001	−0.41
	G	417	0.31	0.95				
Gender	F	960	0.05	0.96	−1.63	1730	0.104	−0.08
	M	772	−0.04	0.93				

3.4. Relationships between Reading Tests and the Language Awareness Test – a correlational study

The *Language Awareness Test* (LAT) was subjected to correlational analysis with the reading aloud tests that were prepared to assess the level of decoding and comprehension. Table 5 offers the results of the relationship between the two variables in terms of decoding and Table 6 in terms of comprehension.

In the theoretical part of this paper, we referred to sources that examined the relationship between language ability and reading skills. One could therefore expect to see higher correlation coefficients in terms of comprehension as opposed to decoding. Yet, as shown in Table 5, these relationships were confirmed, albeit with varying degrees of substantive and statistical significance.

The number of words read in one minute does not show as strong a relationship with the LAT in Grade 1, but from Grade 2 onwards these are correlation coefficients with high statistical significance. We can find reasons for this in the stage of the development of the reading skill itself; reading is only in its early stages at this point and variables other than morpho-syntactic ability may enter into reading performance.

A similar interpretation is possible in relation to other parameters of reading performance – the total time needed to read the text and reading accuracy (the score reflects the total number of words read in one minute and the number of misread words).

We have interesting results for accuracy scores. Here, the significance of the relationships is strongest in the highest age group, i.e. in the fourth- and fifth-graders. In general, the accuracy is high in this age group, so it seems that if the child reads inaccurately, then it will most probably be a child with difficulties with the recognition of morphemes and syntactic structure.

Overall, the child's morphosyntactic abilities, as measured by the LAT, are related to the reading skills affected by our tests. The differences in statistical significance between subtests A and B are not very pronounced.

Table 5: *Relationship between the LAT and decoding*

Grade	Title	Language Awareness Test		
		Total score	Subtest A	Subtest B
	Number of words read correctly in the first minute			
1	Preparing for the Journey	0.25*	0.20*	0.24*
2–3	Feeding Rabbits	0.44***	0.40***	0.41***
4–5	Ice-skating	0.41***	0.30***	0.42***
	Total time taken for the reading			
1	Preparing for the Journey	−0.15	−0.12	−0.15
2–3	Feeding Rabbits	−0.40***	−0.36***	−0.37***
4–5	Ice-skating	−0.34***	−0.24***	−0.37***
	Accuracy			
1	Preparing for the Journey	0.21*	0.20*	0.19*
2–3	Feeding Rabbits	0.07	0.05	0.07
4–5	Ice-skating	0.35***	0.28***	0.34***

Table 6 shows the relationships between the LAT performance and comprehension scores in our new tests. Consistently with the expert literature, we can conclude that there is a relationship between the LAT and the comprehension tests – that is, in their respective modes. The relationships are mostly highly statistically significant ($p < 0.001$ or $p < 0.01$, respectively). While *Listening Comprehension* correlates significantly with the *Language Awareness Test* scores, there were no statistically significant relationships with the test of oral reading for first-graders (*Preparing for the Journey*) and a weaker relationship with the test of silent reading (*The Snowman*). This is consistent with the findings that early in the development of reading skills, reading comprehension is more strongly influenced by the level of decoding (ability to read words) than by language ability.

The *Control Vocabulary Test* showed highly statistically significant relationships with the LAT, which was designed to assess the child's morphosyntactic abilities. This is not surprising as both relate to higher levels of language.

Table 6: *Relationship between the LAT (total score) and comprehension tests (listening, oral reading, silent reading, Pearson's r)*

Grade		Total score	Explicit comprehension	Implicit comprehension	Interpretation	Control Vocabulary Test
Listening comprehension tests						
1	The Forest Elves	0.23**	0.15*	0,21**	0,18*	0,30***
2–3	The Fidgety Little Star	0.56***	0.51***	0.44***	0.38***	0.60***
4–5	How to Mushroom	0.45***	0.36***	0.32***	0.28***	0.44***
Oral reading comprehension tests						
1	Preparing for the Journey	0.04	0.10	0.11	0.10	0.10
2–3	Feeding Rabbits	0.56***	0.44***	0.40***	0.50***	0.42***
4–5	Ice-skating	0.50***	0.37***	0.36***	0.37***	0.52***
Silent reading comprehension tests						
1	The Snowman	0,23*	0.17	0.26**	0.13	0.15
2–3	Big Friends	0,60***	0.51***	0.48***	0,.41***	0.45***
4–5	A Trip to Kořenov	0,48***	0.46***	0.35***	0.18**	0.34***

*Statistically significant values on the following level: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*

4. Discussion and conclusion

A number of international research studies have pointed to the key role of morphological awareness in relation to the development of reading skills (Carlisle 2003; Nagy, Berninger, and Abbott 2006; Nunes and Bryant 2006, etc.) and scholars continue to discuss the inclusion of morphological awareness within models of reading development (e.g. Kirby et al. 2012). One cannot say that the topic has not been under discussion in the Czech Republic, where Žlab addressed it in detail in the 1980s and 1990s (Žlab 1992); however, counselling professionals today lack developmental data and the tools to measure these language abilities. We have therefore built on our initial research on the development of reading comprehension (Kucharská et al. 2015) and continued to develop a new diagnostic tool aimed at mapping morpho-syntactic awareness. In our work we refer to morphosyntactic awareness as the handling of linguistic resources on the basis of prior knowledge of the principles of the structure and functioning of the language system. Out of the possible measuring methods, we have chosen an orally administered format that maps the production abilities of the observed phenomena.

The present study aimed to analyse the *Language Awareness Test* (LAT) results in the *PorTex* test battery. Norms are currently being developed and the results can be used for this purpose.

As for the design of the LAT test itself, it turned out that not all the subtests provided satisfactory results. Subtests A and B show the developmental nature of the items through item analysis, i.e. that children in higher grades perform better than those in lower grades and at the same time we also see different percentages reached in each item, which allows for differentiation in the assessment. Good reliability (Cronbach's alpha, test-retest reliability) was also demonstrated. On the other hand, subtest C does not meet the requirement for developmental performance of the items. In many cases the items performed identically (yet

with high scores) from Grade 1 onwards – the subtest therefore does not have much differentiating power. In addition, low reliability was also found. This may be due to the nature of the answers – three out of four questions were answered by selecting a set of possible answers, while intonation was also at play here. This may account for the higher scores as opposed to if open-ended answers had been required. However, we are not abandoning the Sentence Comprehension subtest and will continue to work on modifying or redesigning the items.

Given the specifics of reading instruction in the Czech Republic, we also sought to answer the question of whether performance on the LAT is related to the method used for teaching reading and gender. If differences were found, it would be necessary to adjust the standards as well. Differences did indeed emerge. Children who are taught using the genetic method (spelling) scored better on the LAT with a medium effect and the preparation of the standards will reflect this finding. Differences between boys and girls have not been demonstrated.

The final finding concerned the LAT's relationship to the newly constructed comprehension tests. According to the theoretical background, the relationships were shown to exist, with text comprehension having higher correlation coefficients than decoding. Nevertheless, language ability as measured by our new test shows a relationship with both variables of interest. This general assessment has two exceptions: the relationships are weak or even non-existent for first-grade pupils. Other factors probably also play a role (ability to work, comprehension of instructions) in listening or oral reading. Also, a weak or no relationship with the LAT was shown for silent reading, again in the lower grades. Here, the child's ability to work may have once again figured in the results. In fact, findings of low correlations between morphological awareness and reading skills in the early years of elementary school are also shown in international studies (Kirby et al. 2012). They cite overly difficult tasks in the morphological awareness test as a possible explanation, as well as the developmental nature of the relationship between morphological awareness and reading skills related to the development of reading for meaning.

On the basis of the data presented, we believe that the LAT demonstrates good psychometric properties, and we consider it an important component of the newly developed *PorTex* literacy skills test battery. We believe that the LAT will be beneficial not only for diagnostics and educational psychologists (counsellors), but also for research.

References

Caravolas, M., and J.Volín. 2005. *Baterie diagnostických testů gramotnostních dovedností pro žáky 2. až 5. ročníků ZŠ*. Praha: IPPP ČR.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., and C. Hulme. 2012. Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science 23*(6): 678–686. https://doi.org/10.1177/0956797611434536.

Carlisle, J. F. 2003. Morphology matters in learning to read: A commentary. *Reading Psychology 24*: 291–332.

Daneš, F., Hlavsa, Z., and M.Grepl. 1987. *Mluvnice češtiny: vysokoškolská učebnice pro studenty filozofických a pedagogických fakult. [Díl 3], Skladba*. Praha: Academia.

Deacon, S. H., and J. R. Kirby. 2004. Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics 25*: 223–238.

Dokulil, M., et al. 1986. *Mluvnice češtiny: vysokoškolská učebnice pro studenty filozofických a pedagogických fakult aprobace český jazyk. [Díl] 1, Fonetika, Fonologie, Morfonologie a morfemika, Tvoření slov*. Praha: Academia.

Foorman, B. R., Petscher, Y., and M. D. Bishop. 2012. The incremental variance of morphological knowledge to reading comprehension in grades 3–10 beyond prior reading comprehension, spelling, and text reading efficiency. *Learning and Individual Differences 22*(6): 792–798. https://dx.doi.org/10.1016/j.lindif.2012.07.009.

Gough, P. B., and W. E. Tunmer. 1986. Decoding, reading and reading disability. *Remedial and Special Education* 7: 6–10.

Grepl, M., et al. 2008. Příruční mluvnice češtiny. Praha: Lidové noviny.

Grepl, M., and P. Karlik. 1998. *Skladba češtiny*. Olomouc: Votobia.

Hájková, E., et al. 2013. *Čeština ve škole 21. století – III.: jazykové jevy v dětských prekonceptech.* Praha: Univerzita Karlova v Praze, Pedagogická fakulta.

Helus, Z. 2012. Reflexe a problémy gramotnosti. Komentář k monotematickému číslu.[Reflections on Problems of Literacy] *Pedagogika. Počáteční gramotnost* 62(1–2): 190–204.

Hirschová, M. 2012. *Syntaktické rozbory: studijní text.* Praha: Univerzita Karlova v Praze, Pedagogická fakulta.

Hoover, W. A., and P. B. Gough. 1990. The simple view of reading. *Reading and Writing: An Interdisciplinary Journal 2*(2): 127–160. https://doi.org/10.1007/BF00401799

Internetová jazyková příručka [online] 2008–2021. Praha: Ústav pro jazyk český AV ČR, v. v. i. Cit. 29. 12. 2021. https://prirucka.ujc.cas.cz/.

Karlík. P. 2017. Sloveský rod. In P. Karlík, M. Nekula, and J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. https://www.czechency.org/slovnik/SLOVESNÝ ROD

Kirby, J. R., Deacon, S. H., Bowers, P. N., et al. 2012. Children's morphological awareness and reading ability. *Reading and Writing* 25: 389–410 (2012). https://doi.org/10.1007/s11145-010-9276-5

Kucharská, A., Seidlová Málková, G., and K.Špačková. 2015. Porozumění čtenému – vývojová dynamika a jeho předpoklady. In *Porozumění čtenému III. Typický vývoj porozumění čtenému – metodologie, výsledky a interpretace výzkumu,* 85–126. Praha: UK PedF.

Kucharská, A. 2014. *Riziko dyslexie: Pregramotnostní schopnosti a dovednosti a rozvoj gramotnosti v rizikových skupinách.* Praha: PedF UK.

Kucharská, A., et al. 2015. *Porozumění čtenému III. Typický vývoj porozumění čtenému – metodologie, výsledky a interpretace výzkumu*. Praha: Pedagogická fakulta Univerzita Karlova v Praze.

Kucharská, A., and Šmejkalová, M. 2017. Jazykové uvědomování u dětí mladšího školního věku. *Gramotnost, pregramotnost a vzdělávání 1*(1): 37–65.

Kucharská, A., et al. 2021. *Klíčové gramotnostní dovednosti u žáků základních škol. Manuál diagnostické testové baterie pro žáky 1. stupně ZŠ: A. Teoretická východiska, popisy testů, administrace, vyhodnocení.* Praha: Pedagogická fakulta Univerzita Karlova v Praze.

Matějček, Z., et al. 1987. *Zkouška čtení*. Bratislava: Psychodiagnostika.

Nagy, W., Berninger, V. W., and R. D. Abbott. 2006. Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology* 98: 134–147.

Nübler, N., Biskup, P., and S. Kresin. 2017. Vid. [Online.] In P. Karlík, M. Nekula, and J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. https://www.czechency.org/slovnik/VID

Nunes, T., and P. Bryant. 2006. *Improving Literacy by Teaching Morphemes*. New York: Routledge.

Snowling, M. J., and C. Hulme. 2012. Annual Research Review: The nature and classification of reading disorders – a commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry, and Allied Disciplines 53*(5): 593–607. http://doi.org/10.1111/j.1469-7610.2011.02495.x

Šmejkalová, M. 2012. Teorie didaktických situací – nová výzva pro didaktiku morfologie českého jazyka. *Didaktické studie* 4(1): 42–63.

Tyler, A., and W. Nagy. 1990. Use of derivational morphology during reading. *Cognition 36*: 17–34.

Ziková, M. 2017. Supletivismus. In P. Karlík, M. Nekula, and J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. https://www.czechency.org/slovnik/SUPLETIVISMUS

Žlab, Z. 1992. *Zkouška jazykového citu.* Ostrava: Microdata.

Appendix

Measures of the *PorTex* test battery

- **Phonological skills**
  - Tests of phoneme awareness
  - Pseudowords repetition task
- **Linguistic awareness test** (assessing morphology, word formation, and sentence comprehension)
- **Decoding tests**
  - Word reading test
  - Pseudoword reading test
- **Comprehension tasks**
  - Listening comprehension tests
  - Oral reading comprehension tests
  - Silent reading comprehension tests
- **Reader's self-concept assessment scal**e
- **Environmental factors questionnaires** (Questionnaire for schools, Questionnaire for parents).

LingBaW *Linguistics Beyond And Within 8 (2022), 133-151*

Ground Zero: A bibliometric analysis of L2 vocabulary research 1986-1990

Paul Meara
Swansea University, United Kingdom
University of Oxford, United Kingdom

Abstract
This paper uses a co-citation analysis to examine the research on vocabulary acquisition that was published in 1990. Two analyses are presented. The first is a detailed account of the 1990 research on its own terms. The second analysis places this work in a larger context by looking at research published in a five-year window covering 1986-1990.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric, author co-citation

1. Introduction

This paper is the tenth in a series of studies in which I have been mapping out the way L2 vocabulary research has developed over the last 50 years. Beginning with 1982, *LingBaW* has published a set of papers in which I have presented bibliometric mappings of the research that appeared in each year up until 1989 (Meara 2014–2021). This paper takes this historical overview another step forwards. It covers the research published in 1990 and recorded in the Vocabulary Acquisition Research Group Archive (VARGA. Meara: n.d.), a very large collection of papers that is not limited to the obvious English language sources.

The paper falls into two parts. Part 1 reviews the 1990 research in its own terms. Part 2 puts this research into a wider context by summarising the main trends that appear in a five-year window covering 1986-90. Both parts use the author co-citation method, developed by Small (1973). Small's method is described in detail in Appendix 1 for readers who are not yet familiar with the approach used in these papers.

2. Part 1. The new research published in 1990

At first sight, 1990 does not look like a good year for L2 vocabulary research. 1989 *was* a fairly good year in terms of the number and variety of papers that appeared. In 1990, however, we

have a return to a lower level of activity. The VARGA database (Meara n.d.) identified 142 relevant sources published in 1989. The equivalent figure for 1990 is only 113 published sources – the lowest total since 1985. The obvious interpretation of this decline is that the surge in research that began in the early 1980s has now peaked. However, this interpretation turns out to be somewhat over-simplified. The main reason for this is that 1990 actually sees a significant shift in the type of work being published. A large proportion of the 1989 output appears as book chapters – with the collection by Tickoo (1989) alone accounting for a significant fraction of the total output. There are fewer special issues of journals and edited volumes in 1990, and as a result both the number of book chapters and the number of ordinary papers has declined in this year's output (cf. Figure 1).

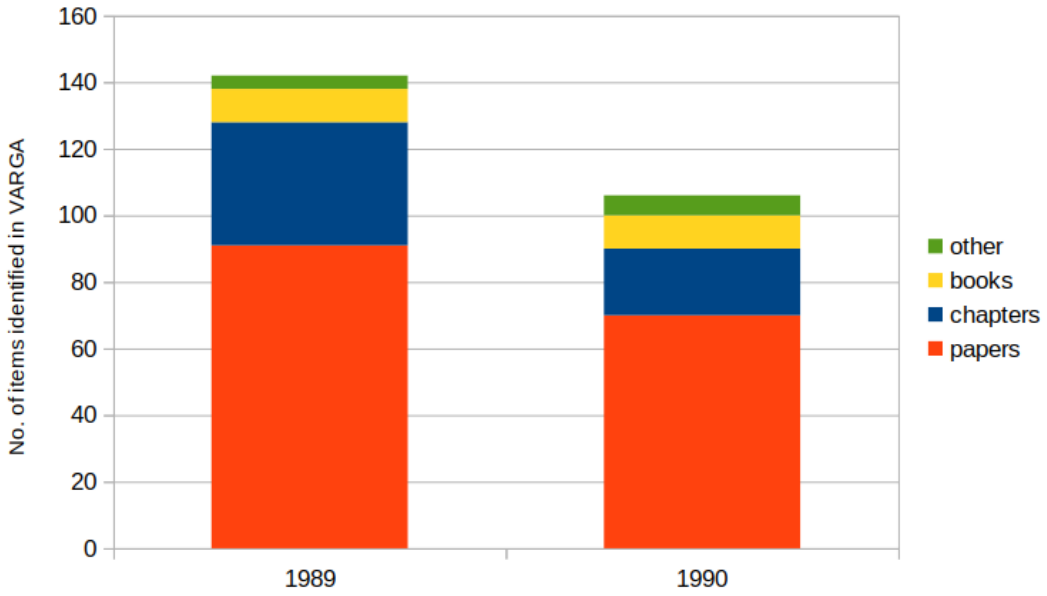


Figure 1: *The 1990 research by item type*

The outstanding publishing event of 1990 is the appearance of Paul Nation's book *Teaching and Learning Vocabulary* (Nation 1990). As we shall see in further papers in this series, it is difficult to overstate how influential this book would turn out to be. Basically it defined a research agenda that came to dominate the field for the next two decades and beyond. Nation's book was not universally welcomed at the time – partly because it shifted the research away from the dominant linguistic areas of research to more psychological aspects of vocabulary – reading, inferencing, learning, testing – which had played only a limited part in the growth of vocabulary research in the previous few years. It is not often that a single book revolutionises a field, but this book is definitely one of them, and we will see its influence coming to dominate the field in later papers in this series.

However, Nation's work was not the only book length treatment to appear in 1990. McCarthy's *Vocabulary* (McCarthy 1990) is a more traditional volume, that focusses more on the linguistic features of vocabulary. The book falls into two halves. Section one (Explaining vocabulary) deals with word-formation, lexical relations, prototypical vocabulary, vocabulary in use, and measurable characteristics of words in texts. These are all features that rely heavily on the corpus research that we noted in the 1989 research. Section two (Demonstrating

vocabulary) is more concerned with vocabulary teaching and the role of the teacher in this process. It covers vocabulary selection, organising vocabulary in a non-random way, presenting vocabulary in classrooms, vocabulary learning strategies and lexical reference (dictionaries and similar support material).

Taylor's book, also called *Teaching and Learning Vocabulary* (Taylor 1990), covers much the same ground as the second half of McCarthy's text. A general introduction to words is followed by chapters that cover communicative vocabulary teaching, repetition and interaction, and exercises for consolidating newly learned vocabulary. The book ends with a short chapter on vocabulary in discourse.

A radical departure was Willis' *The Lexical Syllabus* (Willis 1990). Like McCarthy (1990), this book also leans very heavily on research carried out by the COBUILD team at Birmingham University. It argues that most EFL text books pay too much attention to grammar, and not enough attention to words, and suggests that much of the grammatical structure of English can be found by studying the behaviour of words in a corpus. This was not just a question of emphasis: Willis argues that it radically changes the roles of teachers and syllabus designers, and that it encourages learners to become active explorers of a language, rather than passive recipients of their teachers' wisdom.

Ljung's book *A study of TEFL vocabulary* (Ljung 1990) also relies heavily on a corpus based approach to vocabulary, though unlike the three previously mentioned texts, it is mainly concerned with a specific set of textbooks. The book falls into two parts: Part One is a detailed critical analysis of the vocabulary that appears in a set of 56 English language textbooks aimed at secondary school pupils in Swedish schools. This analysis is actually quite short – only 40 pages. Part Two – just short of 400 pages contains detailed word lists that inform the discussion in Part One. This is an unusually thorough analysis of the vocabulary taught in textbooks, and it deserves to be much better known than it appears to be. (At the time of writing, *Google Scholar* lists fewer than 100 citations of this work, compared to 348 for Taylor, 1242 for Willis and 2191 for McCarthy.)

A number of other titles are worth mentioning briefly. Dretzke and Nester (1990), Tréville (1990) and Milan and Sunkel (1990) rehash some familiar discussions about cognates and false friends.

Diab (1990) is a longer version of a study that was first reported in 1989. It looks at how nurses access ESP vocabulary by using dictionaries. Murphey (1990) analyses the lyrics of pop songs, and discusses how they can be used to teach words to learners of English. Schrameier (1990) and Willems and Oud-de Glas (1990) were unobtainable due to COVID restrictions in place at the time this work was being carried out.

Four cited theses dealing with L2 vocabulary acquisition were published in 1990. These are listed in Table 1.

Table 1: *Theses published in 1990 that are cited in later years*

Altarriba, J
Constraints on interlingual facilitation effects in priming in Spanish-English bilinguals. PhD Dissertation. Vanderbilt University. 1990.
Cunningham, L
L2 vocabulary: A study of the word association responses of beginning learners of Irish. MPhil Dissertation. University of Dublin. 1990.
de Jong, E
Woorden leren met prentenboeken. Een vergelijking van twee instructiemethodes. PhD thesis. Katholieke Universiteit Brabant.– Tilburg. 1990.
Dekkers, R
Woorden schieten tekort. Een onderzoek naar de optimalisering van woordenschatverwerving bij allochtone kinderen. PhD thesis. Katholieke Universiteit Brabant. Tilburg, 1990.

2.1. The data sources

As usual in these reports, the analysis presented in this section is based on papers published in journals or as book chapters. The monographs and theses listed in the previous section are not included in the analysis as these longer works tend to use citations in ways which differ from what we find in shorter research papers.

The VARGA data base (Meara: n.d.) lists a total of 95 sources eligible for the analysis, but a surprisingly large number of these items were not available – partly because of COVID travel restrictions and library closures at the time I was collecting this data. These items – 17 in total – were excluded from the analysis, but they are listed here in Table 2.

It is unlikely that the exclusion of these items significantly affects the analysis that follows, but a number of points are worth noting. Firstly, several of these items are German sources, and their appearance in Table 2 reflects the ongoing library issues arising from the political events in Germany in 1989. Secondly, the paper by McCarthy was one of several short think-pieces that appeared in a thematic volume of *English Studies* published by the British Council. These papers are best described as and somewhat eclectic, in their coverage, but they probably reflect a growing interest in vocabulary in the British Council, an influential player in English Language Teaching at the time. Thirdly, Hayes (1990) is the first paper dealing with acquisition of Chinese that we have identified in this series of analyses. Fourthly, the omission of Lado is probably the most serious omission. Lado wrote a series of important vocabulary research papers in the1960s and 1970s where he looked at "massive vocabulary expansion". At the time vocabulary acquisition was not a significant topic of research and experimental studies of vocabulary acquisition were not fashionable with the result that this work largely went unnoticed. (Google Scholar indicates that Lado's *Final Report* is cited only two times.) Finally, the omission of Zimmermans' paper in Milwaukee Studies significantly reduces the importance of this German strand of research.

The remaining 78 studies are not listed here for space reasons, but interested readers can find the complete list on the VARGA web-site: *https://www.lognostics.co.uk/varga/*. Entering **1990 ##** in the search box will return a complete list of the papers that are included in this year's analysis. Although the number of eligible papers is considerably lower than the figure of

118 papers that we identified in 1989, it is actually close to the number of papers and book chapters published in 1988 (83). This suggests that the relatively large size of the 1989 data set might be an unusual blip in the numbers, rather than an indicator of a reliable upward trend in the number of outputs.

Table 2: *Eligible items published in 1990, but not included in the formal analysis*

Abe, H. 1990. The acquisition of communicative vocabulary: A case of procedural knowledge. *FL Reporter,* 22:33-45.

Anon. 1990. Wortsschatzarbeit = Vokabellernen? *Neusprachliche Mitteilungen,* 43:101-103.

Appel, R. and A. Vermeer. 1990. Woordenkennis sleutel voor schoolsucces: uitbreiding woordenschat in kleuterklassen. *Stimulans,* 8,7:6-8.

Fanning, H. 1990. Falsche Freunde in zweisprachiger Erziehung (Englisch/Deutsch). In: H. J. Zobel (ed.) *Studien zur Sprachdifferenzierung: Roland Arnold zum 60. Geburtstag.* Wissenschaftliche Beiträge der Ernst-Moritz-Arndt-Universität Greifswald. 47-51.

Farrell, P. 1990. Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary. *CLCS Occasional Paper* No. 25. Trinity College Dublin.

Hayes, E. 1990. The relationship between 'word length' and memorability among L2 readers of Chinese Mandarin. *Journal of Chinese Language Teachers Association,* 25,3:31-41.

Hinz, K. 1990. Wortschatzarbeit in der leitungschwachen Englischklasse. *Der fremdsprachliche Unterritcht,* 23: 22-26.

Kühn, P. 1990. Das Grundwortschatzwörterbuch. In: FJ Hausmann, O Reichmann, HE Wiegand and L Zgusta (eds.) *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie, 2.* Teil-band. Berlin. 1353-1364.

Lado, R. 1990. Towards a lexico-semantic theory of language and language learning. *The Georgetown Journal of Language and Linguistics,* 1,1: 96-100.

Lipczuk, R. 1990. Über ein Wörterbuch der faux amis. In: A Kątny (ed.) *Deutsche Sprache im Kontrast und im Kontakt.* Rzeszów. 225-237.

McCarthy, M. J. 1990. Minding your words: two important areas of vocabulary learning. *English Studies,* 5:6-7.

Quetz, J. 1990. Wortschatzarbeit. *Zielsprache Englisch,* 20:29-31.

Reiner, E. 1990. "False friends" und "doublets" JA Pfeffers *Germanismenwörterbuch der englischen Sprache. Moderne Sprachen,* 34,1-2:41-48.

Scherfer, P. 1990. Vom Nutzen des Vorwissens im Vokabelunterricht. *Der fremdsprachliche Unterricht*, 24:30-36.

Vater, B. 1990. Schuler verlangen nach erfullter Gegenwart! Wortschatzarbeit in einem handlungsorientierten Englischunterricht in der Sekundarstufe 1. *Der fremdsprachliche Unterricht,* 23: 11-16.

Zimmerman, R. 1990. Productive and unproductive lexical strategies in L2 learners. *Milwaukee Studies on Language*. 1990.

We begin by reporting the usual superficial analysis of authorship. A total of 94 authors make a contribution to the 1990 output. Again, this number is considerably down on the 1989 figures, where 134 unique authors were identified, but close to the figure that we reported for 1988 (83 unique authors). In 1990, the proportion of authors contributing to just one paper is 88%. Table 3 shows the distribution of authors in 1989 and 1990 in terms of the number of contributions that they make to the two data sets.

Table 3: *The number of authors contributing to N outputs in the 1989 and 1990 data sets*

Contributions	**6**	**5**	**4**	**3**	**2**	**1**
1990 data			2	0	6	87
1989 data			1	4	18	109

The best interpretation of these figures seems to be that that the proportion of authors contributing to only one paper has risen slightly in the 1990 data set (85% compared with 82% in 1989).

The number of authors contributing to two or three papers has fallen dramatically, but there is a small rise in the number of authors contributing to 4 papers in this data set. Laufer and Meara both contribute to four papers; Appel, Broeder, Colpaert, Decoo, Schouten-van Parreren and Swartz contribute to two papers each. It is worth noting that the first five of these authors all worked in the Netherlands, underlining the importance of vocabulary research in that country at the time. Swartz is a scholar working on CALL at the US Army Research Institute, Alexandria. Of the authors who contributed two or more papers in 1989, most have fallen out of the major contributors list for 1990. Of the 1990 authors, only Meara and Laufer were also significant contributors to the 1989 output and all the authors of two papers are new entrants into this list. This is a surprisingly high level of churn, and it suggests that the field as a whole is still a long way from settling down into a steady pattern of outputs.

2.2. The analysis

The main analysis reported in this section is not directly concerned with who published in 1990, but rather with the sources that are cited in the 1990 data set. Cited authors provide us with clues as to the ideas that were important at the time, and allow us to plot the growth of research trends. The first step in our analysis, then, is to identify the important sources that are cited in the 1990 data set. The methodology for doing this has been described in detail in the previous papers in this series (see Appendix 1). For 1990, the analysis identifies 1363 sources – very close to the number we identified in the 1988 data set (1391), but considerably fewer than we identified in the larger 1989 data set (1911). As usual, most of these sources are cited only once, but a small number of sources are cited much more than this. The data is summarised in Table 4.

Table 4: *The number of times sources are cited in the 1990*

frequency	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**
cases							1					
frequency	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**
cases			1	3	2	4	5	5	18	25	55	168

The most cited sources in this data set are Meara (19 citations), Nation (11 citations), Carter, Cohen and Krashen (ten citations each), Richards and Laufer (nine citations each), Palmberg, McCarthy, Lockhart, and Levenston (eight citations each). Most of the sources that appear in this list are new: only Nation, Meara, Carter and Levenston were also highly cited in the 1989 data. Direct comparisons between the 1990 data set and the 1989 data set are difficult because the 1990 dataset is a lot smaller than that of the previous year. Inclusion in the ten most cited list for 1989 required a source to be cited at least 11 times, whereas the equivalent threshold for 1990 was only eight citations. Meara's 19 citations (22% of the total data set) is striking, as it exceeds the previous "record" set by Nation in 1989.

The next step in our analysis of the 1990 data is to identify the most frequently cited sources in the data set. Once these sources have been identified, we can construct a map based on the co-citations between them. The convention here is to identify a set of about 100 sources, but for 1990 it is difficult to do this. For the 1989 data set, we adopted an inclusion threshold of 5 citations, and this gave us a set of 80 sources to map. With the 1990 data, adopting this threshold would leave us with only 43 sources to map. Lowering the threshold to four citations adds a further 25 sources, bringing the total to 68 cases, and this is as close as we can get to the 1989 figures. In percentage terms, a source that is cited four times in this data set is cited in nearly 5% of the papers published in 1990.

The co-citation data for the 68 most cited sources in the 1990 data set were extracted from the complete data set (all 1363 sources cited in 1990), and the results analysed using the Gephi software package (Bastian, Heymann and Jacomy 2009). Gephi's analysis of these co-citations is shown in Figure 2. The interpretation of this map is relatively straightforward. Gephi identifies seven clusters in this data set.

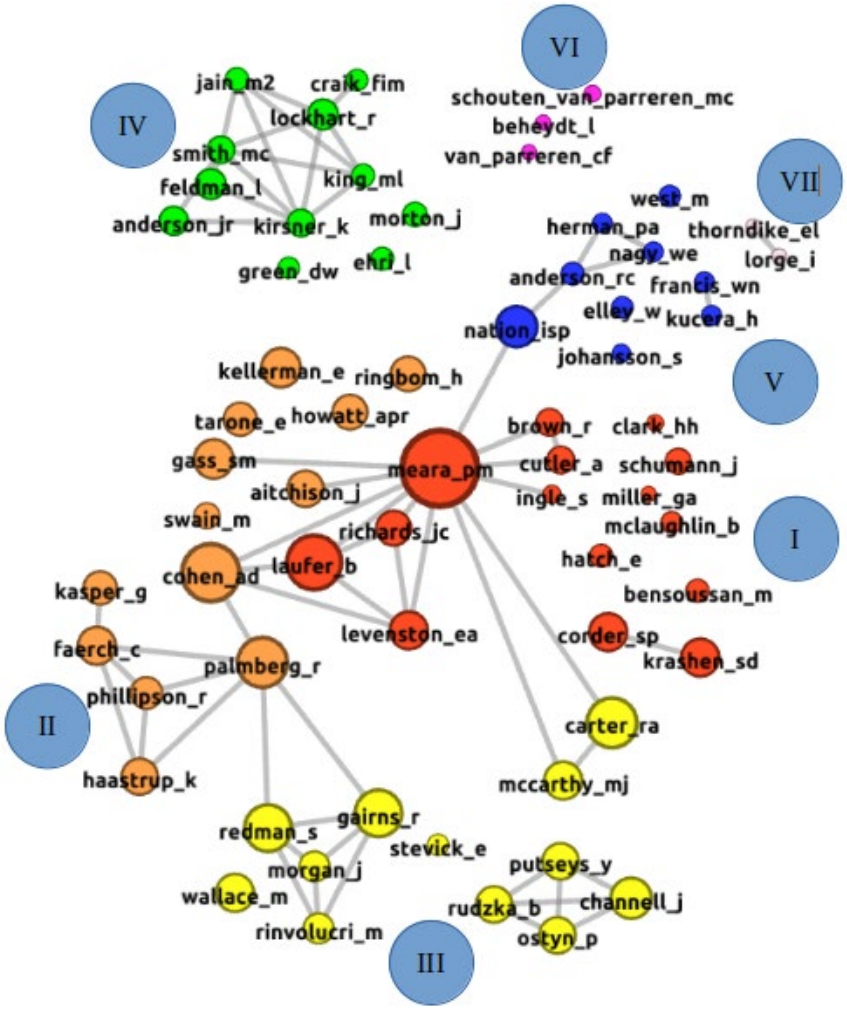


Figure 2: *A co-citation map of the 1990 data set. The map contains 68 nodes, each representing a source that is cited at least four times in the data set. The number of citations is reflected in the size of the node. Edges between nodes indicate that the two nodes are co-cited at least four times. The nodes are grouped into clusters taking into account all co-citation links that appear at least twice in the complete data set. Weaker edges have been removed from the map in the interest of simplicity and clarity.*

The core of the map is made up of **Cluster I** (15 sources, centred on Laufer, Meara, Richards and Levenston). These sources are perhaps best described as the current mainstream of L2 vocabulary research, but the sub-cluster made up of Corder and Krashen is influential too. The detached sources in this cluster signal psychological influences on the L2 vocabulary research. Brown is an important figure in L1 acquisition; Hatch and Clark work in the area of semantics; Schumann represents a strand of neurolinguistic research; Miller is best known in this context for his work on the limitations of short term memory; Cutler is best known for her work on slips of the tongue; Bensoussan is closely associated with Laufer; McLaughlin (1987) is a psychologically oriented text book dealing with L2 acquisition.

Cluster II is a group of mainly European researchers who are loosely interested in the psycholinguistics of L2 vocabulary acquisition. The core of this cluster is a group of Scandinavian researchers that we have identified in our earlier analyses (Faerch and Kasper, Haastrup, Phillipson, Palmberg and Ringbom). Andrew Cohen is the most cited source in this cluster, mainly cited for his interest in mnemonic approaches to vocabulary acquisition (e.g. Cohen 1987).

Cluster III is a loose group of sources who are mainly concerned with vocabulary pedagogy. We can identify three sub-clusters here. McCarthy and Carter represent input from corpora and discourse analysis (e.g. Carter and McCarthy 1988); Rudzka, Ostyn, Putseys and Channel authored a series of influential textbooks that focussed on semantic relations between words (e.g. Rudzka et al. 1985); Stevick (1976) is an important methodological source; Gairns and Redman (1986), Wallace(1982) and Morgan and Rinvolucri (1986) are methodology textbooks mainly aimed at teachers. With the exception of Stevick, all the sources in this cluster are UK based.

Cluster IV is the now familiar group of cognitive psychologists who work with experimental studies of bilinguals. This cluster is detached from the other clusters in the map. The sources here are not really interested in the pedagogical aspects of L2 vocabulary acquisition, though some of this work – notably Craik and Lockhart's (1972) paper, and Craik and Tulving's 1975 paper – are strongly cited in the L2 literature.

Cluster V, centred on Nation, is principally concerned with reading in an L2. Nagy, Herman and Anderson are important L1 reading theorists. Kucera and Francis, Johansson and West are word lists.

Cluster VI is a group of Dutch speaking vocabulary researchers.

Cluster VII (Thorndike and Lorge) is another word frequency count, and is probably best treated as an extension of the word counts in Cluster V.

Most of these clusters will be familiar from the maps published earlier in this series of papers.

Figure 3 shows the changes that have taken place between the 1989 data set and the 1990 data set.

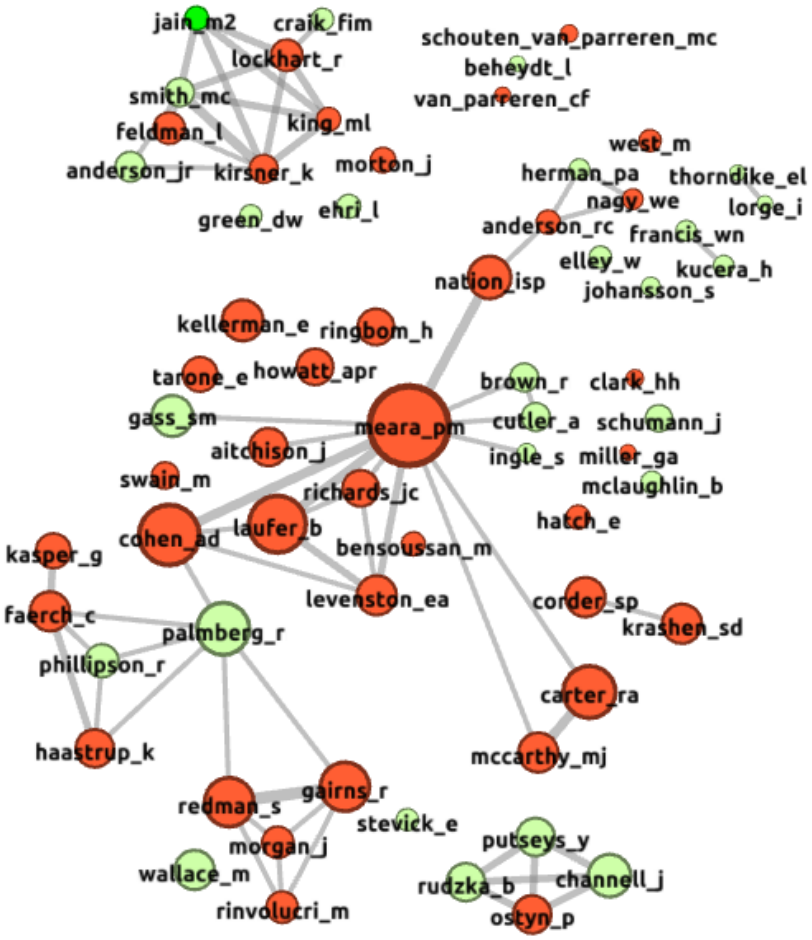


Figure 3: *New entries in the 1990 data set (shown in green)*

The figure preserves the cluster layout of the 1990 data set, but uses colour to show the appearance of new sources. The 1990 "survivors" – sources that appear in both the 1989 and 1990 data sets are shown with dark shading. There are 35 survivors in 1990, about 54% of the total, indicating a fair degree of stability in this data set. The 30 new cases – sources who did not appear in the 1989 data set – are shown with light shading. However, many of these apparently new entries had already appeared in our earlier maps, and are best described as "returners" rather than new entries. Only seven genuinely new entries appear in the 1990 data set: Cutler, Ingle, Schumann and McLaughlin in Cluster I, Phillipson in Cluster II, Stevick in Cluster III and Green in Cluster IV. Clusters V, VI and VII consist entirely of familiar sources. Cutler represents an important strand of L1 research on speech errors. Schumann is mainly cited for his work on neurolinguistics. McLaughlin's textbook, as we have already noted, is more psychologically oriented than other works of this type. These new sources may indicate that Cluster I is becoming more oriented towards the psycholinguistics of L2 vocabulary development, while the more traditional linguistic influences are moving to other clusters, and perhaps becoming less influential. Phillipson, in Cluster II, is working closely at this time with Haastrup, and represents a significant strengthening of the Scandinavian sub-cluster that we have noted in previous years. Stevick, in Cluster III, is a methodological textbook. Green, new in Cluster IV, signals an additional research theme among the psycholinguistic sources.

Green is mainly cited for his work on the way bilinguals control access to their two languages (e.g. Green 1986).

The most striking feature of the 1990 data set, one which is not immediately apparent in the maps, is that two of the clusters that made up the 1989 map have almost completely disappeared in 1990. The most important cluster in 1989 was a densely connected dictionaries/semantics cluster. Surprisingly, all the members of this cluster fail to appear in the 1990 map. The implications of this change are not entirely clear. It is possible that the dictionary theme has played itself out by 1990, but this interpretation will need to be confirmed in later maps. The 1989 map also contained a small cluster whose focus was best described as L2 vocabulary learning in German and French. This cluster too has disappeared from the 1990 map. A small L1 acquisition cluster from 1989 has been reduced to a single node in 1990 (Clark in Cluster I).

We also find some smaller changes in the 1990 map. There is a suggestion that vocabulary course books (cluster III) are becoming more coherent as a feature, but less strongly connected to the empirical and theoretical work thinking that characterises Clusters I and II. There is also a suggestion that the psycholinguistics cluster has grown in 1990, and is more diverse in its membership than it was in 1989. It is not clear whether this is a genuine shift, or just a reflection of different authorship patterns in psycholinguistics – papers in this area tend to have multiple authors, who all cite each other, and this can sometimes inflate the importance of a source. Cluster VI, the Dutch/Flemish cluster, appears in the 1990 map as a detached cluster, no longer widely co-cited by the key sources in Cluster I and Cluster V.

3. Part 2. A wider perspective 1986-90

In this section, we will place the 1990 data into a larger context by looking at a five-year window covering all the research published between 1986 and 1990. Working with a five year window smooths out some of the short-term fluctuations in the annual data, and the results are less obviously affected by the size of the individual data sets.

For the purposes of comparison, Table 5 summarises the main features of the 1985-1989 data.

Table 5: *The main characteristics of the 1985-1989 data set*

Number of papers in the data set	477
Number of authors contributing to the data set	475
Number of sources cited in the data set	4616
Inclusion threshold for this data set	14 citations
Number of cited sources meeting the inclusion threshold	103
Identifiable co-citation clusters	6+1 detached singleton
I: vocabulary acquisition (26)	
II: dictionaries and corpus analysis (25)	
III: reading meaning and inferencing (23)	
IV: word recognition in an L2, performance of bilingual speakers (19)	
V: mental imagery (5)	
VI: Français Fondmental (5)	

Table 6, summarises the main characteristics of the 1986-1990 data set. Here, the 1985 data has fallen out of the five year window, and the 1990 data has been added to the window. Because 1985 was a (relatively) good year for L2 vocabulary research, and 1990 saw a slight fall in the number of publications, the 1986-1990 data set is marginally smaller than the 1985-1989 data set that we reported in last year's analysis, and the number of authors contributing to the 1986-90 data set is a lot smaller. In other respects the two data sets are comparable.

Table 6: *The main characteristics of the 1986-1990 data set*

Number of papers in the data set	465
Number of authors contributing to the data set	411
Number of sources cited in the data set	4699
Inclusion threshold for this data set	14 citations
Number of cited sources meeting the inclusion threshold	99
Identifiable co-citation clusters	6
I: vocabulary acquisition (41)	
II: dictionaries, corpus analysis and semantics (25)	
III: L2 word recognition, performance of bilinguals, imagery (20)	
IV: Reading and word frequency counts (7)	
V: Dutch vocabulary research (4)	
VI: French vocabulary research (1)	

Table 7 shows the number of contributions each of the 411 authors made to the 1986-1990 data set.

Table 7: *The number of authors contributing to N papers in the 1986-1990 data set, and the expected number of authors based on Lotka's Law*

Papers	**19**	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Authors	1				1		1	1	2		2	6	9	16	53	335
Lotka	0	1	2	2	2	3	3	4	5	7	9	13	21	37	84	335

Once again, we find that the vast majority (82%) of the authors contribute to only one paper. This figure is slightly larger than the equivalent figure for the 1985-1989 window. However, the number of authors contributing to multiple publications still falls short of what we would expect to find. 14 authors contributed to five or more publications in the 1986-1990 window. Thirteen of these authors were already identified as prolific authors in the 1985-89 data set. Here, their status appears to be consolidated: Meara (19 papers), Zimmerman (12 papers), Laufer (10 papers), Palmberg (9 papers), Broeder and Carter (8 papers each), Beheydt (6 papers), AD Cohen, Extra, Nation, Robinson, van Hout and Vermeer (5 papers each). Schouten-van Parreren (6 papers) is the only new addition to the most prolific author list. Two prolific authors from the 1985-89 have fallen off the list in the 1986-1990 window: McCarthy contributes to 4 papers in the 1986-90 window, and Alfes contributes to two papers in this data set.

The bottom line of Table 7 shows the number of authors we would expect to contribute to N publications in a data set of this size. These figures are based on a suggestion by Lotka (1926). Lotka's method is outlined in Appendix 2 for readers who are not already familiar

with this approach. The table shows that, given 335 authors who contribute to just one output, we might expect 84 authors who contribute to two outputs, 37 authors who contribute to three outputs, and so on. The table shows that the actual figures fall quite a long way short of Lotka's projection for this data set. Meara is the only author who bucks the general trend at this time.

Table 8 summarises the number of times each source is cited in the 1986-90 data set. As usual, most sources are cited only once in the complete set of 465 papers – 2967 authors, or 63% of the total of 4699 sources fall into this category. However, there is some evidence that the number of sources who are very frequently cited is increasing: the most frequently cited sources are Meara (cited in 72 papers), Nation (60), Krashen (51), Levenston (50), Faerch and Richards (46), Carter and Sinclair (37), Kasper (36), Kellerman, Lockhart and Schouten-van Parreren (35 citations each). Most of these names will be familiar from our earlier analyses. This list is largely identical to the list of highly cited sources in the 1985-89 data set, but four new sources appear in the 1986-90 data set – Carter, Sinclair, Kasper and Lockhart – replacing four sources who have fallen out of the 1985-89 data set – AD Cohen, West, Blum-Kulka and Corder.

Table 8: *The number of times sources are cited in the 1986-90 data set*

FREQUENCY	**60+**	**59**	**58**	**57**	**56**	**55**	**54**	**53**	**52**	**51**	**50**	**49**	**48**	**47**	**46**
CASES	2									1	1				2
FREQUENCY	**45**	**44**	**43**	**42**	**41**	**40**	**39**	**38**	**37**	**36**	**35**	**34**	**33**	**32**	**31**
CASES									2	1	3	1	2	1	
FREQUENCY	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**
CASES	2	1		3	1	2	4	6	3	8	7	7	6	10	6
FREQUENCY	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
CASES	5	10	14	26	24	25	35	41	44	89	113	181	327	673	2967

Next, we turn to an analysis of the co-citation patterns among the most frequently cited sources. Standard practice in this type of analysis is to work with the 100 most frequently cited sources, and Table 8 shows that we can use a threshold of 14 citations to identify a suitable set of 99 sources for the analysis that follows. This threshold is identical to the threshold we adopted in last year's analysis.

Data for these 99 sources was analysed using the Gephi package. The results of Gephi's analysis are shown in Figure 4, which maps the 99 most frequently cited sources in the 1986-90 data set, using 14 citations as a threshold for inclusion. The sources are linked by 575 edges each of which occurs at least six times. However, in the interests of simplicity, Figure 4 shows only those edges which occur eight or more times. Gephi finds six clusters in this data set.

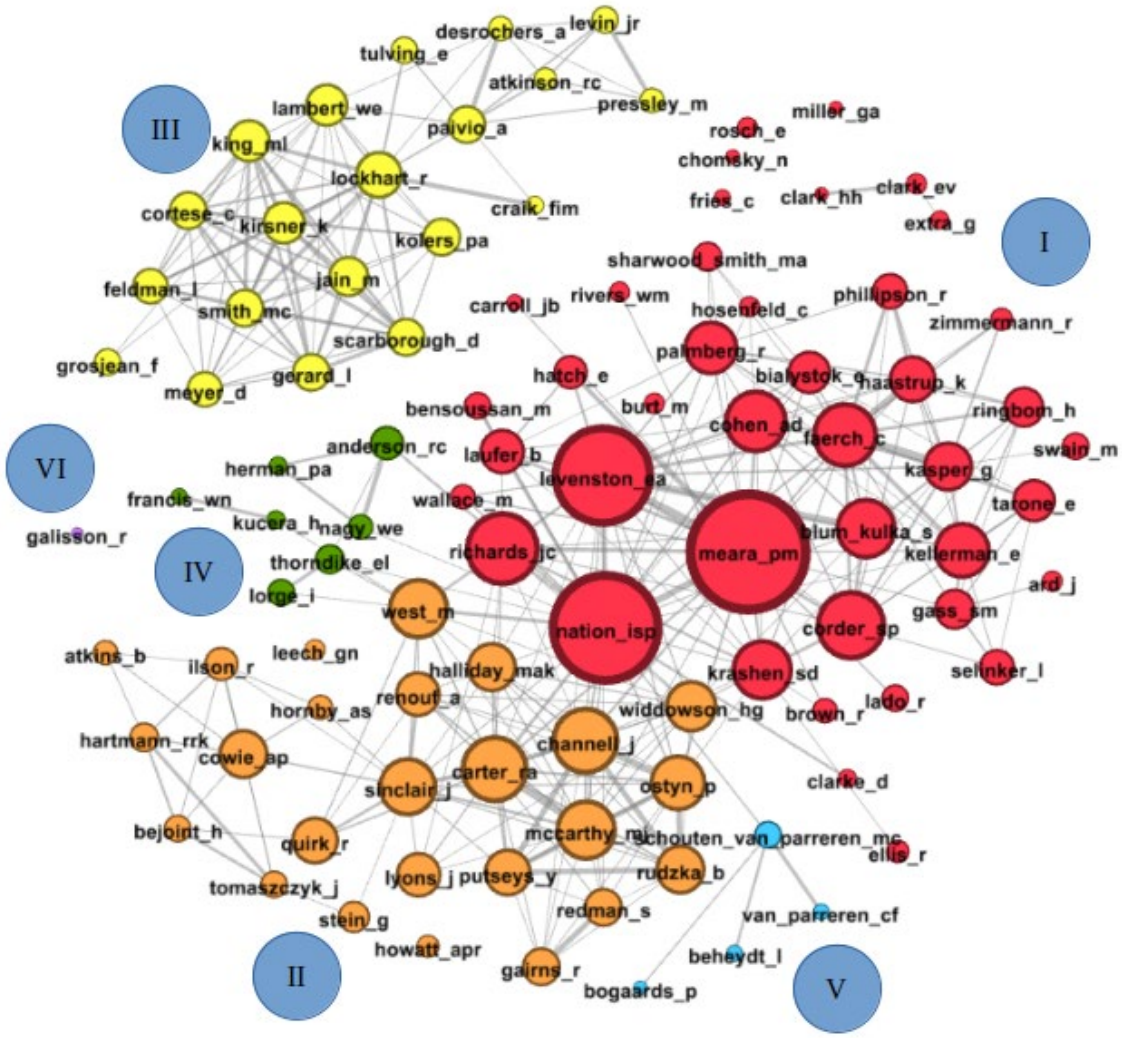


Figure 4: *A map of the co-citations found in the 1986-1990 data set. Nodes are sources that are cited at least fourteen times in the data set. Co-citation links that occur less than eight times in the data set have been eliminated in the interests of simplicity.*

The four most cited sources in the dataset (Meara, Nation, Levenston and JC Richards) form the core of **Cluster I.** This cluster is best described as the mainstream L2 vocabulary cluster at this time. It consists of 36 sources that are strongly co-cited, a small detached subcluster that is clearly a set of L1 acquisition sources (H Clark, E Clark and Extra) and a small number of detached outliers (GA Miller, cited most often for his work on memory chunking; Rosch, cited most often for her work on semantics; Fries, cited most often for his work on semantic frames; and Noam Chomsky as the goto reference for generative grammar.) A number of other sub-clusters can also be identified here. The most important of these is the Scandinavian group (Faerch, Kasper, Haastrup, Phillipson, Ringbom and Palmberg). We have noted the significance of this group in our earlier maps, but by 1990 this group is becoming a key feature of the L2 vocabulary research enterprise. The Faerch/Kasper co-citation, for example, is by far the strongest co-citation in this data set (34 occurrences). We can also identify an influential Israeli sub-cluster (Levenston, Laufer, AD Cohen, Blum-Kulka and Bensoussan), a small Edinburgh sub-cluster (Corder, Selinker), and a subcluster consisting of

Krashen and Burt. These last two sub-clusters were important in previous maps, but their influence here seems to be in decline.

Cluster II is the now familiar Dictionaries and Semantics cluster that has also appeared in our earlier maps (24 sources + 1 detached outlier, Leech). Three sub-clusters can be identified here, one dealing specifically with dictionaries (Hartmann, Stein, Tomaszczyk, Cowie), a second that deals more broadly with corpora-based approaches to vocabulary (Quirk, Sinclair, Carter, McCarthy and Halliday) and a third subcluster consisting of important textbooks (Gairns and Redman, Channel, Ostyn, Rudzka and Putseys). A cluster of this type was clearly identified in our analysis of the 1985-1989 data: here it has shrunk slightly. The distinguishing characteristic of this cluster is that its members are co-cited with Nation and JC Richards, but only rarely with the other sources who make up Cluster I.

Cluster III is the familiar group of psychologists and psycholinguists whose work influences L2 vocabulary research (20 sources). At first glance, this cluster appears to be about the same size as the psycholinguistics group we identified in last year's map. However, its composition has changed a bit. Sources that were previously important in this cluster, notably Lambert and Tulving, now seem to be less so. A couple of sources have dropped out of the cluster (Albert and Obler, Macnamara, and Caramazza), and their place has been taken by a group of sources who work on imagery that formed an independent cluster in the 1985-1989 map. The cluster continues to be dominated by a team led by Kirsner.

Cluster IV, also familiar from previous years, is mainly composed of frequency counts. It contains a subcluster of L1 reading specialists who are strongly co-cited with Nation, but not with other members of Cluster I.

Cluster V is a group of L2 vocabulary researchers who publish mainly in Dutch. The key figure in this group is Schouten-van Parreren who is often co-cited with Nation, but only infrequently co-cited with other sources in Cluster I.

Galisson stands as the sole member of **Cluster VI.** He is the one remaining representative of French vocabulary research in this map, but his work is not well-integrated with that of the other sources.

4. Discussion

Sharp-eyed readers will have noticed that the map in Figure 4 looks rather different from the five-year maps in our previous analyses, which were becoming increasingly complex and difficult to interpret. Gephi offers a number of different map lay-outs, and Figure 4 is computed using the Fruchterman-Reingold algorithm, which places well-connected nodes at the centre of the map, and less well-connected nodes towards the periphery. Hopefully, use of this format will allow future analyses to be compared more easily with earlier maps, by introducing a greater degree of standardisation to the maps. For example, the Fruchterman-Reingold layout allows the size of the maps to be standardised in a way that was not possible with the layouts that we have been using up till now. Furthermore, it is possible that this format will make it easier to carry out some further analysis that takes greater account of the strength of the connections between the more important nodes. These features will be

exploited in future analyses. (See https://www.lognostics.co.uk/maps/ for a further discussion of this issue.)

Meanwhile, a number of themes are beginning to emerge from this analysis of the 1986-1990 dataset. Firstly, and most importantly, the main L2 vocabulary cluster has grown substantially, mainly by absorbing what appears as a separate cluster in the 1985-1989 map. The cluster has almost doubled in size by 1990, and now accounts for about half of the nodes in the map. The members of this cluster are very strongly cited alongside each other, and almost all of the members of the cluster can themselves be identified as active L2 vocabulary researchers. I think this is a sign that the L2 vocabulary research is, by this time, becoming self-sufficient, and less reliant on external influences. The few external influences that remain are clearly becoming less central to the structure of this cluster: L1 vocabulary acquisition has almost become detached (Clark and Clark); error analysis (Corder and Selinker) is clearly marginal; Krashen's influence on the field also seems to be on the decline. At the same time, this cluster has started to become isolated from the rest of the map. This is not immediately apparent from Figure 3, but a closer look at the data suggests that Nation, Levenston, Meara and Richards are the only sources who have strong connections outside the cluster. The implications of this drift are not entirely clear.

The second theme to note concerns Cluster II, the dictionaries, corpora and semantics cluster. The membership of this cluster is almost unchanged in the five year map. It is worth remembering, however, that this theme no longer has a strong presence in the one-year 1990 map, so the strength of this theme might be expected to diminish in future years. Meanwhile, Whitcut has dropped out of the cluster, Kucera and Francis have moved to another cluster, and the cluster has gained three new sources in their place (Gairns and Redman, and Howatt). The cluster as a whole is notably self-contained: members of this cluster are co-cited with the main sources in Cluster I, but there are few other strong co-citations. This level of stability in a cluster is unusual, and it strengthens the view that research in this area may have peaked.

The third theme emerging from this map concerns the role of national groupings. The main feature here is that the Français Fondamental cluster no longer figures in the 1986-90 map. Robert Galisson (Cluster VI) remains as the sole representative of this important French research strand (though he is actually quite critical of the approach that the Français Fondamental project took). Likewise, the German research that we noted in our earlier maps is again represented only by Zimmerman, who plays a minor role in Cluster I. Dutch vocabulary research has become detached from the main L2 research cluster, and now stands alone as Cluster V in the 1986-90 map. The increasing dominance of English language research – both in the sense of research written in English and research about learning English vocabulary – is very clear here. Other languages barely get a look in.

Finally, the main fault line in the 1986-90 map lies between Cluster III and the other clusters. The map suggests that the co-citation links between sources in Cluster III and the other sources in the map are non-existent. In reality, there are some weak links, but they are not strong enough to show up here. The map also suggests that the composition of this cluster is stable: the cluster contains 19 members in our 1985-1989 map, and 20 in the 1986-1990 map. However, a number of features of the new Cluster III are worth commenting on. The key figure in this cluster is Paivio, whose dual-coding approach to bilingual memory provides

a unifying theme for work on mnemonics and bilingual storage. The mnemonics and imagery sub-cluster that is built around Paivio is largely based on research that was carried out in the late 70s and early 80s. This work has consistently failed to find a home in mainstream L2 vocabulary research, but it shares little with the rest of the psycholinguistic cluster other than an emphasis on experimental methods. At the moment, the position of this sub-cluster looks precarious, and it would not be surprising to find that it disappears from our maps in the future. The second sub-cluster in Cluster III seems to be more robust. Kirsner's group (King, Jain, Lockhart, Smith) and Scarborough's group (Scarborough, Cortese, Gerard) both use lexical decision tasks as a way of investigating bilinguals' mental lexicons. Most of the citations to these sub-clusters appear in the latter half of the 1985-90 window, so it looks as though they will persist into the future. In contrast to Paivio, Grosjean, whose work at this time is largely concerned with code-switching, looks rather isolated at the edge of Cluster III. He is frequently cited alongside the other members of this cluster, but only rarely cited by them.

Notably missing from Cluster III are Albert and Obler. Their 1978 edited collection, *The Bilingual Brain,* (Albert and Obler 1978) was an important influence throughout the early part of the 1980s, but it has only one citation in 1990 (Channel 1990) suggesting that interest in neurolinguistic research on how bilinguals handle L2 words is rapidly falling off.

Overall, the 1986-90 map shows a very high degree of stability. Ninety sources that appeared in the 1985-89 map also appear in the new map; only nine new sources make an appearance. Five of these – Ard, Burt, R Ellis, Fries and Sharwood-Smith – are additions to Cluster I. Two of the new sources appear in Cluster IV (Herman and Lorge); Howatt is a new addition to Cluster II; Bogaards is an addition to the Dutch research group (Cluster V).

5. Conclusion

The analyses in this paper suggest that the consolidation we noted in earlier reports is continuing into 1990. There are some small changes in the prolific author lists, and the relative importance of the major sources, but these changes do not amount to a serious change of direction. Nation, Meara, Levenston and Richards are clearly an important set of pivotal nodes in the maps, as they provide most of the links between the main clusters. Most of the major players in L2 vocabulary research are already in place in these maps, and we can clearly see a sort of consensus beginning to emerge. Meara (2020b) has argued that these developments can be seen as a "First Paradigm" for L2 Vocabulary Research.

Looking forward, we might ask: what research themes will take the place vacated by the dictionary theme? will the French research continue its decline? will the Dutch language group achieve a breakthrough? will the growing Scandinavian research group and the strong Israeli sub-cluster start to have a major influence on the main vocabulary cluster? We might also note that some sources that rise to prominence in later maps, still play only a small role here. How quickly will they develop their full potential? And will their increasing importance create new research themes that we can observe through these maps?

Finally, we might expect that Nation's 1990 textbook will start to emerge as the key source for vocabulary researchers in the very near future, and we might expect to see a big increase in

the number of research papers which pick up the agenda that this book inspired. The next paper in this series will examine the research published in 1991 in the context of a five-year window covering 1987-1991. For the moment, the likely shifts in L2 vocabulary research remain just around the corner.

References

Albert, M. and L. K. Obler. 1978. *The bilingual brain: neuropsychological and neurolinquistic aspects of bilingualism*. New York.

Bastian, M., Heymann, S., and M. Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Carter, R. and M. J. McCarthy (eds.) 1988. *Vocabulary and Language Teaching*. London.

Channell, J. 1990. Vocabulary acquisition and the mental lexicon. In J. Tomaszczyk and B. Lewandowska-Tomaszczyk (eds.) *Meaning and lexicography,* 21-31. Amsterdam.

Cohen, A. D. 1987. The use of verbal and imagery mnemonics in second language vocabulary learning. *Studies in Second Language Acquisition* 9: 43-62.

Craik, F. I. M. and R. S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11: 671-684.

Craik, F. I. M. and E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104: 268-294.

da Solla Price, D. 1965. Networks of scientific papers. *Science* 49(3683): 510-51

Diab, T. 1990. *Pedagogical Lexicography: A case study of Arab nurses as dictionary users.* Berlin.

Dretzke, B. and M. I. Nester. 1990. *Student's guide to false friends, new friends and old friends.* Berlin.

Gairns, R. and S. Redman. 1986. *Working with words: a guide to teaching and learning vocabulary*. Cambridge.

Green, D. W. 1986. Control, activation and resource: a framework for the control of speech in bilinguals. *Brain and Language* 27: 210-223.

Ljung, M. 1990. *A study of TEFL vocabulary*. Stockholm.

Lotka, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12): 317-324.

McCarthy, M. 1990. *Vocabulary*. Oxford.

McLaughlin, B. 1987. *Theories of second-language learning.* London.

Meara, P. M. 2012. The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal* 43(1): 7–22.

Meara, P. M. 2014. Life before Nation: Bibliometrics and L2 vocabulary studies in 1982. In M. Gómez González, F. J. Ruiz de Mendoza Ibañez, F Gonzálvez García and A. Downing (eds.), *The Functional Perspective on Language and Discourse: Applications and implications,* 111-129. Amsterdam.

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. 2020a. The British are coming...! A bibliometric analysis of L2 vocabulary research in 1989. *Linguistics Beyond And Within* 6: 111-129.

Meara, P. M. 2020b. The emergence of a First Paradigm in vocabulary research: The bibliometrics of *System*. *Vocabulary Learning and Instruction* 9(1): 1-32.

Meara, P. M. 2021. A time for consolidation: A bibliometric analysis of L2 vocabulary research 1985-89. *Linguistics Beyond and Within* 7: 65-85.

Meara, P. M. (no date) VARGA: The Vocabulary Acquisition Research Group Archive. Available at: https://www.lognostics.co.uk/varga/ (accessed 25th October 2022).

Milan, C. and R. Sünkel. 1990. *Falsche Freunde auf der Lauer: Dizionario di false analogie e ambigue affinità fra tedesco e italiano*. Bologna.

Morgan, J. and M. Rinvolucri. 1986. *Vocabulary*. Oxford.

Murphey, T. 1990. *Song and music in language learning*. Bern.

Nation, I. S. P. 1990. Teaching and Learning Vocabulary. Rowley, Mass.

Rudzka, B., Channell, J., Putseys, Y. and P. Ostyn. 1985. *More Words You Need*. London.

Schrameier, A. 1990. *Wortbedeutung im Gedächtnis*. Göttingen.

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.

Stevick, E. 1976. *Memory, meaning, and method*. Rowley, Mass.

Taylor, L. 1990. *Teaching and learning vocabulary*. New York.

Tickoo, M. L. (ed.) 1989. *Learners' dictionaries. State of the Art*. Singapore.

Tréville, M-C. 1990. *Good friends: the use of English cognates for learning French*. Quebec.

Wallace, M. 1982. *Teaching Vocabulary*. London.

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.

Willems, M. M. and M. M. B. Oud-de Glas. 1990. *Vocabulaire selectie voor het vreemdetalenonderwijs*. Nijmegen: ITS.

Willis, D. 1990. *The Lexical Syllabus*. London.

Acknowledgements

Thanks to Marlene Schwarz for help with sourcing the German literature. Special thanks to John Read for his insightful comments on the earlier papers in this series.

Appendix 1: 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. Self-citations, where an author cites their own work, are treated in the same way as any other citation, on the grounds that authors only rarely fail to cite their own work. 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.

Appendix 2: 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	63	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 1986-1990 data set when x= 2.7, 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.

LingBaW *Linguistics Beyond And Within 8 (2022), 152-166*

Domains of case changing and case maintaining movements

Mark Newson and Krisztina Szécsényi
Eötvös Loránd University, Budapest, Hungary

Abstract
With recent developments in Case Theory, movements in which a DP acquires a different case to the one it would have received had it not moved have been accepted as a possibility. In this paper we examine a number of such movements from a variety of languages to attempt to characterise and understand them more fully. Based in Dependent Case Theory, our analysis claims that case change does not really happen, but case assignment is allowed to be delayed under certain circumstances creating the illusion of one case over-writing another. In explicating these circumstances, we are not only able to provide a better understanding of when 'case change' can and can't happen, but also develop the theory in ways which address certain conceptual problems that it faces.

Keywords: Dependent Case Theory, domains, unmarked case, case change

1. Introduction

Early developments in Minimalist syntax suggesting that case features were not responsible for the surface distribution of DPs opened up new paths for investigating case phenomena with numerous positive consequences. One of these was that empirical observations from a number of languages seemingly showing movement from one 'case position' to another could be properly investigated rather than being dismissed as being nothing to do with case (see Szabolcsi 1994, for example) or just flatly ignored.

In this paper we investigate such case changing movements and attempt to account for the difference between case changing and case maintaining movements. In the next section we will introduce our adopted framework, Dependent Case Theory, and its development by Baker (2015). We then investigate a number of case changing movements to try to establish differences between them and case maintaining movements. After showing that there is no one factor which accounts for the differences by itself, we argue that it is a combination of factors which define the conditions in which case change under movement is possible. An account of why these particular factors in combination allow for case change is proposed and the repercussions of this for the framework investigated.

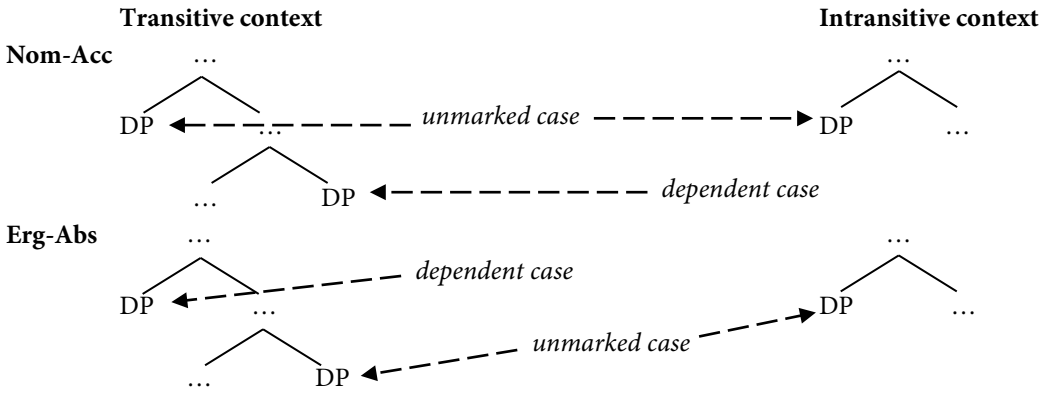
2. Dependent Case Theory

Dependent Case Theory was first proposed by Marantz (1991) and subsequently developed by Baker (2015). Marantz claimed that different types of cases are assigned to a set of eligible DPs in a given order: lexical (inherent) case > dependent case > unmarked case > default case. As lexical and default case will not enter our analysis, we will discuss only dependent and unmarked case here.

Dependent case is dependent on the structural configuration for its assignment. It can only be assigned in transitive contexts; that is, there must be at least two items on the list of case eligible DPs. If this condition is met, dependent case can be assigned to one or the other of the two, giving rise to different case systems, as we will demonstrate shortly. Once a DP is assigned case, it drops off the list. Any DPs remaining on this list are then assigned unmarked case.

Depending on which DP is assigned the dependent case, either a nominative-accusative system or an ergative-absolutive system will result, as shown in the following diagram:

(1)



Note that the single DP in the intransitive context always gets unmarked case, as in this configuration dependent case cannot be assigned. Thus, either the object (absolutive) or the subject (nominative) of the transitive context gets the same (unmarked) case as the subject of the intransitive context.

3. Domains

In Marantz's system, the set of DPs on the list for any determination of case assignment was kept local to each other via the standard Case Theory notion of government by the same head. One of the developments that Baker (2015) introduced was to make the system more in-line with Minimalist assumptions by limiting the set of eligible DPs through a notion of a case domain. These domains, Baker asserted, are essentially the phases relevant for structure building (Chomsky 2001). Case assignment should be seen as a part of the Spell Out process and hence takes place when a phase head is introduced. In this way, only those DPs introduced into the structure since the last spell out would be considered for case assignment.

Another advantage of the notion of a domain is that different dependent and unmarked cases can be defined with respect to separate domains of a language. So, while nominative

might be the unmarked case of a clausal domain, genitive could be the unmarked domain of the nominal domain; ergative might be the dependent case of the clause, but inside the VP dative might be dependent. This allows a language to have more than just two structural cases. However, although Baker's use of domains draws case theory closer to the rest of the minimalist system, it highlights an unexpected difference between case assignment and other syntactic operations, which he does not discuss. For movement purposes, *v* is assumed to be a universal phase head, making VP a fixed Spell-Out domain. However, it appears that languages differ in terms of whether VP counts as a case domain or not. In some languages the presence of an object within the VP has a role in determining the assignment of dependent case at the clausal level, Thus, both the subject and the object must be visible at clause spell out. This implies that VP is not a case domain. This is so for Hungarian, for which it is usually assumed that an object to the right of the verb remains inside the VP and one to the left occupies a VP external position:[1]

(1)

a. *János* [VP *kivágta* *a* *fá-t*].
John down-chop-PAST-3S the tree-ACC
"John chopped down the tree."

b. *János* *a* *fát* [VP *vágta* *ki* -].[2]
John the tree-ACC chop-PAST-3S down
"It was the tree that John chopped down."

Regardless of where the object is placed, however, it is assigned dependent (accusative) case, which means that it is evaluated for case assignment in the same domain as the subject and hence the VP is not a domain for case assignment. In Sakha, however, it makes a difference whether the object is within the VP or not. As Baker points out, the indefinite DP remains inside the VP while the definite DP is extracted out of it:

(2)

a. *Masha türgennik* [VP *salamaat sie-te*].
Masha quickly porridge eat-PAST.3SS
'Masha ate porridge quickly.'

b. *Masha salamaat-y* *türgennik* [VP – *sie-te*].
Masha porridge-ACC quickly eat-PAST.3SS
'Masha ate the porridge quickly.'

As we can see, the VP internal indefinite object is unmarked for case, whereas the extracted definite is accusative. Baker interprets this under the assumption that in this language VP is a case domain, so the VP internal object is evaluated independently of the subject. As it is the only DP in the domain, it is assigned unmarked case. The extracted DP, however, is moved

[1] We use the following abbreviations in examples: cases (small caps) NOM = nominative, ACC = accusative, GEN = genitive, DAT = dative; tenses (small caps) PAST = past, PRES = present, INF = infinitive, AOR = aorist; persons are given by numbers 1, 2, 3; number (small caps) S = singular, PL = plural; grammatical function (capitals) S = subject; miscellaneous (small caps) POSS = possessor, DEC = declarative, AF = affirmative, NOML = nominaliser.

[2] It is possible that the verb has moved out of the VP in this case and only the pre-verb is left inside it. For the purposes of the argument, however, it is only important to recognise that the object is not within the VP.

into the same domain as the subject and therefore is evaluated along with it. As this domain has a transitive configuration, dependent case can be assigned to the object and hence this appears in the accusative.

Note that the two examples above exemplify case maintaining and case changing movements respectively. We will consider further examples in the following section to try to get a broader perspective on the phenomenon.

To account for this typological distinction, Baker introduces two different types of case domain. Phase heads are claimed to be parameterized in terms of the Spell-Out domains they introduce: a *hard phase head* introduces a Spell-Out domain whose contents are fully spelled out immediately and hence the DPs it contains are not available for consideration in subsequent phases; a *soft phase head* introduces a domain whose contents are spelled out, but its DPs remain visible at the next phase and so can play a role in determining the conditions of dependent case assignment, and even be the recipient of such a case.

This may solve the empirical problem, but it raises several conceptual problems in its wake. Merely describing the difference between the two types of domains does not explain how the difference is possible under general assumptions. It clearly does not account for how DPs which have already been spelled out can remain active and even acquire new features in subsequent phases. Moreover, we are also left without explanation of the difference between case and other grammatical processes: why is the notion of a soft domain relevant only to case assignment? We will return to these issues once we have a clearer idea of what determines when movements allow case change and when they do not.

4. Further case changing and case maintaining movements

There are languages which allow case change of a DP extracted from the possessor position. For example, Baker (2015) claims that the following Japanese sentences can be analysed as demonstrating the case change of a possessor, from genitive to nominative, when it is extracted from the possessive DP into a focus position:

(3)

a. [$_{DP}$ *John-no otoosan*]-*ga sin-da.*
John-GEN father-NOM die-PAST
'John's father died.'

b. *John-ga* [$_{DP}$ – *otoosan*]-*ga sin-da.*
John-NOM father-NOM die-PAST
'It is John whose father died.'

There is an interesting contrast between these data and similar phenomena in Hungarian, where an extracted dative possessor maintains its case:

(4)

a. *El-veszett* [$_{DP}$ t_i [*a fiú-nak*] *a kalap-ja*].
away-lost the boy-DAT the hat-3S.POSS
'The boy's hat got lost.'

b. [*A fiú-nak*]$_i$ *veszett el* [$_{DP}$ t$_i$ *a kalap-ja*].
the boy-DAT lost away the hat-3SG.POSS
'It was the boy's hat that got lost.'

These observations clearly show that we cannot attribute case change to the kind of movement involved, as presumably a very similar process is used in both (3) and (4).

Furthermore, it doesn't seem that we can attribute the feature involved in case changing movement to specific languages, as although possessor extraction in Hungarian is a case maintaining movement, movement of the possessor within the DP is case changing:

(5)
a. *a fiú kalap-ja*
the boy.NOM hat-3S.POSS
'the boy's hat'
b. *a fiú-nak a kalap-ja*
the boy-DAT the hat-3S.POSS
'the boy's hat'

It is generally accepted (following Szabolcsi 1983, 1994) that the nominative possessor occupies a lower position in the Hungarian DP and that the dative possessor is raised to a higher DP internal position, as indicated by its position relative to the determiner in (5b). As this movement involves case change, one cannot maintain that Hungarian does not allow case changing movement at all. Generally, then, it seems that different movements in different languages can change or maintain case.

5. Possible accounts of the difference

5.1. Unmarked case

Baker (2015) notes that it is only DPs originally assigned unmarked case that experience case change under movement. He provides a number of examples of dependent case marked DPs which do not change their case when moved. For example, in Sakha, which we have seen has case changing movement of definite objects, Baker argues that dative is the dependent case of the VP domain, assigned to the higher DP in double object and causative constructions.[3] The dative case is maintained, however, in constructions in which this DP is moved to subject position, such as the possessive construction in (6):

(6) *Ucuutal-ga student-nar tiij-bet-ter.*
teacher-DAT student-PL suffice-NEG.AOR-3S
'The teacher doesn't have enough students.'

Other evidence in favour of Baker's claim comes from possessor extraction data. We saw that Japanese possessors change case when extracted. Baker argues that genitive case is unmarked

[3] It is because the dative case is not assigned to an argument with a particular thematic interpretation that Baker argues it is not an instance of inherent case, as it is in some other languages.

in this language. Evidence for this is that Japanese assigns genitive to both possessors and noun complements in the DP:

(7) *yuubokumin no toshi no hakai*
nomad GEN city GEN destruction
'the nomad's destruction of the city'

This means that genitive cannot be a dependent case in the nominal domain, as otherwise it would be assigned to either the possessor or the noun complement, but not both. Baker then claims that languages which do not have such double genitive constructions, assign dependent genitive to their possessors and this correlates with case maintaining possessor extraction, as is demonstrated in the Cuzco Quechuan data below:[4]

(8)
a * *Marsila-q Maduna-q llimp'i-sqa-n.*
Marsila-GEN Madona-GEN paint-NOML-3PL
'Marsila's picture of Madona.'
b *Pi-qpa-man-mi qulqi-ta* [– *ususi-n-man*] *qu-ni?*
who-GEN-DAT-AF money-ACC daughter-3-DAT give-1SS
'I gave money to whose daughter?'

Despite these observations, there is reason to believe that Baker's claim that DPs with unmarked case are subject to case change on movement cannot be the whole story as there are examples where a DP with unmarked case maintains this when moved. Thus, while it may be true that having an unmarked case is a necessary condition for case change, it is not a sufficient condition. Hungarian dative possessor extraction provides an example of an unmarked case which is maintained on movement. Newson and Szécsényi (2020) argue that in Hungarian dative case is the unmarked case of domains associated with a non-finite inflection, such as is found in the possessive DP. One reason to believe this is the fact that this inflection appears on a type of infinitive and such clauses have dative subjects:

(9) *Muszáj* [*Péter-nek haza-men-ni-e*].
must Peter-DAT home-go-INF-3S
'Peter must go home.'

The dative subject of the inflected infinitive behaves in exactly the same way as the nominative subject of the finite clause, in that it appears in both intransitive (9) and transitive contexts:

(10) *Nem szabad* [*Péter-nek meg-néz-ni-e ez-t a film-et*].
not allowed Peter-DAT pv-watch-INF-3S this-ACC the film-ACC
'Peter is not allowed to watch this film.'

[4] Obviously, the fact that the moved wh-phrase in (8b) gets an extra dative case stacked on its original genitive adds and extra complication, which we will not go into here. The important observation, however, is that the extracted possessor maintains the genitive case which was assigned in the possessive DP.

As it follows that dative must be seen as the unmarked case in this domain and as the appearance of the same inflection and a dative DP in the possessive DP seems more than a coincidence, we extend this analysis to dative possessors too. Therefore, the data in (4) provide us with an example of a DP with an unmarked case which is maintained when it is moved out of the possessive DP.[5]

5.2. Soft domains

If it is not the type of case borne by a DP that determines whether or not the DP can change its case when it moves, then perhaps it is the kind of domain that it moves out of that influences the phenomenon. There is reason to believe that movement out of a hard domain does not trigger case change. As DPs do not interact with each other for case purposes across clause boundaries, it seems that TP is a universally hard domain. Movements out of TP, such as wh-movement, never seem to involve case change and this might be taken to suggest that it is only movement out of soft domains which allows DPs to acquire another case in their landing site.[6]

However, once again, while movement out of a soft domain appears to be a necessary condition for case change, it is not a sufficient one as there are examples of such movements which maintain the original case of the DP. We pointed out above that the Hungarian VP must be taken to be a soft domain to allow VP internal objects to interact with the subject in the TP domain. We also showed that extraction of the object from the VP does not result in case change (see (1)). Therefore, not all movements out of soft domains involve changing case.

5.3. The combined approach

So far, we have argued that all case changing movements involve DPs assigned an unmarked case and are movements out of soft domains, but case change is not attested in every moved DP with an unmarked case or every DP extracted from a soft domain. Here we will argue that it is the combination of these two factors which provides the sufficient and necessary conditions for case changing movement.

---

[5] A reviewer suggests that we might bolster the claim that dative is unmarked in Hungarian by considering passive constructions; the passive subject is generally seen as receiving unmarked case. Unfortunately, Hungarian does not have passive constructions. There are other observations, however, which point to the same conclusion. For example, Hungarian allows inherent dative ('quirky') subjects and it is well known that such subjects trigger unmarked case on the object. In finite clauses, verbs with dative subjects do indeed take nominative objects, but in inflected infinitives such verbs have both subjects and objects in dative:

(i) *Mindenkinek tetsz-ik ez a könyv.*
everybody-DAT like.3S this the book.NOM
'Everybody likes this book.'

(ii) *Nem kell mindenki-nek tetsze-ni-e en-nek a könyv-nek.*
Not must everybody-DAT like-INF-3S this-SAT the book-DAT
'It is not necessary for everyone to like this book.'

[6] The Sakha data concerning the movement of definite objects would appear to be a counter example to this if we accept Baker's analysis. We will address this issue towards the end of the paper.

We can use the Hungarian data to set this up. The facts are provided in (5) and (4b), repeated here as (11) for convenience:

(11)

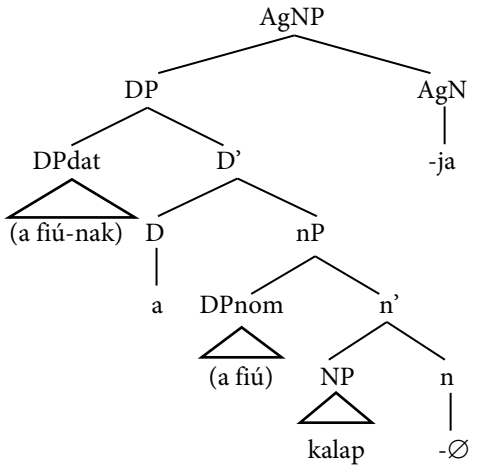
a. *a fiú kalap-ja*
the boy.NOM hat-3S.POSS
'the boy's hat'

b. *a fiú-nak a kalap-ja*
the boy-DAT the hat-3S.POSS
'the boy's hat'

c. [*A fiú-nak*]$_i$ *veszett el* [$_{DP}$ t$_i$ *a kalap-ja*].
the boy-DAT lost away the hat-3S.POSS
'It was the boy's hat that got lost.'

The lower possessor position is a nominative position, whereas the higher one is dative. Only extraction from the dative position is possible, which indicates that both the complement of the determiner and the DP are phases.[7] Newson and Szécsényi (2020) argued that this is because both the determiner and the non-finite agreement are phase heads, assuming the following structure:

(12)



Being a phase head, the determiner induces the spell out of its complement, headed by *n*, the element which introduces the possessor. The possessor is the only DP in the case domain and hence is eligible for unmarked (nominative) case. We claim that the Hungarian determiner is a soft phase head and hence nP is a soft case domain. Consequently, movement of the possessor out of this domain allows for case change, both conditions for this being met. Once the possessor has moved to the specifier of DP, it finds itself in a different case domain. This time, the phase head is the non-finite agreement (AgN), which is a hard phase head. The

[7] A reviewer asks about wh-movement in Hungarian as a way of supporting our claims that nP is a soft domain and DP a hard one. In fact, wh-movement works in exactly the same way as focus movement does: an extracted wh-possessor is always dative and therefore has to move via the DP specifier. However, a wh-possessor can never be nominative and therefore must move to the DP specifier position whether extracted further or not. Because of this, we do not get a complete paradigm of all possibilities for possessors, which is why we concentrated on focus movement here.

possessor, however, is still the only one in the domain and hence is again eligible for unmarked case, which is dative for this domain (as evidenced by the dative subject of the inflected infinitive in (9) and (10)). If the possessor moves from this position, as in (11c), case change will not happen as the domain it moves from is a hard one.

The claim that the DP is a hard domain is supported by the fact that the phase head, AgN, also introduces clauses, which are always hard domains. Further evidence comes from the comparison of the Hungarian possessive DP to the equivalent structure in Korean. As we saw in (3), the Japanese possessor changes its case when extracted from the possessive DP. From our perspective that must mean that the Japanese DP differs from the Hungarian one in being a soft rather than a hard domain. Nakamura (2002) demonstrates that, like Japanese, Korean possessors can receive nominative case, but in Korean they can also receive accusative:

(13)

a. *Swunsang-uy* *pyeng-i* *choykun* *simha-ta.*
The Prime Minister-GEN illness-NOM recently serious-PRES

b. *Swunsang-i* *choykun* *peyng-i* *simha-ta.*
The Prime Minister-NOM recently illness-NOM serious-PRES
'The Prime Minister is seriously ill.'

(14)

a. *Mary-ka* *John-uy* *tali-lul* *cha-ss-ta.*
Mary-NOM John-GEN leg-ACC kick-PAST-DEC

b. *Mary-ka* *John-ul* *tali-lul* *cha-ess-ta.*
Mary-NOM John-ACC leg-ACC kick-PAST-DEC
'Mary kicked John's leg.'

This is somewhat problematic for Baker's account of double nominative constructions, which assumes that the extracted possessor is an adjunct and as such is evaluated after the other arguments have been case marked. This means that the possessor is evaluated as the sole remaining DP in the domain, and hence is assigned nominative. Presumably the same analysis is applicable to (14b), but then it is unexpected that the possessor would get accusative rather than nominative. Furthermore, as Ryu (2014) points out, the label 'double' nominative/ accusative construction is a misnomer as the assignment of nominative and accusative case is not restricted to two DPs:

(15)

a. [*ttokki-ka*] [*kwi-ka*] [*kkuth-i*] [*thel-i*] *kil-ta.*
rabbit-NOM ear-NOM top-NOM fur-NOM be.long-DEC
'The fur of the top of the ears of the rabbit is long.'

b. *Hans-ka* [*ttokki-lul*] [*kwi-lul*] [*kkuth-ul*] *cap-ass-ta.*
Hans-NOM rabbit-ACC ear-ACC top-ACC grab-PAST-DEC
'Hans grabbed the top of the ears of rabbits.'

While Baker's extraction analysis might remain valid for one of the nominative possessors, as the word order in (13) suggests, it is unlikely that it can be applied to all the other cases. Whatever the mechanism that is in operation in these multiple nominative and accusative constructions, the case system must have access to DP internal DPs at the clausal level. It

therefore follows that the DP must be a soft domain in this language. No such similar observations can be made in Hungarian and we argue that this is precisely because the Hungarian DP is a hard domain

6. Emerging questions

The data discussed so far strongly indicate that we are on the right track when we propose a combined account according to which a DP assigned unmarked case in its extraction site can be assigned another case in its landing site if it moves out of a soft domain.

There are several questions that emerge from this though, which, once sufficiently answered, can lead to deeper generalizations, and therefore lend greater explanatory power to the theory. We will address three of these:

1. Why does the combination of the initial assignment of unmarked case and movement out of a soft domain result in case change?
2. How is case change even possible if it involves the movement of a DP from one Spell-Out domain to another?
3. Why is the soft/hard distinction relevant for case phenomena but not for movement?

This last point is especially pertinent given that the problem arises from Baker's attempt to unify movement and case through the identification of case domains as Spell-Out domains.

In Baker (2015) the distinction between hard and soft domains is a mechanism needed to overcome a problem without much explanatory content. Our proposal will refine the definition of the soft domain, leading to an elevation Baker's distinction to a more explanatory level.

In order to do so, we reach back to Marantz's (1991) original idea that once a DP has been assigned a case, it drops out of the competition and has no further role to play, either to receive another case or to determine the case assigned to another DP. We will term this assumption the Invisibility Principle:

(16) **The Invisibility Principle 1**
Once a DP is assigned case it becomes invisible to the case system.

Clearly Baker's assumption that a DP with unmarked case is available for further case marking goes against the Invisibility Principle.[8] But we can recast the idea to extend Baker's attempt to incorporate Case Theory into Phase Theory:

(17) **The Invisibility Principle 2**
Once a DP is *fully* spelled out, it is inactive in higher phases.

[8] Baker marginalises this part of Marantz's system by parameterising it. He claims that there are some ergative languages which still assign dependent case to the subject when the object is assigned an inherent case and so the object must still be visible after it is case marked. He rejects the possibility that ergative might be an inherent case as it is not tied to a specific interpretation. However, this does not rule out the possibility that ergative is a 'lexical' case, in the sense of Woolford (2006).

Lying behind this definition is the idea that a DP will only be fully spelled out when it is assigned a case. What needs to be addressed is when exactly case features are assigned. This is where the distinction between soft and hard domains turns out to play a central role. According to the Invisibility Principle those DPs which seem to undergo 'case change' cannot actually be assigned case before they move and hence must still be active in the case system. This is exactly what the soft/hard domain difference boils down to in our proposal which defines soft domains as follows:

(18) **Soft domain**
A soft domain is one for which unmarked case is not assigned at its point of spell-out.

Hard domains in turn are defined as domains where all cases are assigned at spell-out. The effect of this is that the assignment of unmarked case within a soft domain will be delayed until the next hard domain. Therefore, a DP eligible for unmarked case in a soft domain remains active until the next hard domain and, if it moves, it will be eligible for the case determined for it in its landing site. In other words, 'case change' never really happens, but some DPs can be eligible for different cases, depending on their position at the point of spell out of a hard domain.

The Invisibility Principle also offers a principled account of why the distinction between hard and soft domains is relevant for case but not for movement. What keeps a DP active after the domain it is generated in is spelled out is its lack of case and case features play no role in determining whether or not movement takes place. Therefore, the distinction between hard and soft domains will be relevant only for case assignment.

7. Analysis

We are now in a position to offer an account for why the movement of some DPs 'changes' the case of those DPs while the movement of others does not. In this section we present some sample derivations to illustrate our claim.

The central component of our proposal is the assumption that unmarked case assignment in soft domains is delayed until the next hard domain. It has the following empirical consequences: (i) DPs eligible for unmarked case in soft domains remain visible after the domain has been spelled out; (ii) these DPs are assigned case as determined in the subsequent hard domain, potentially resulting in 'case change'; (iii) unmarked case will also be fixed in hard domains and so movement out of hard domains is case maintaining.

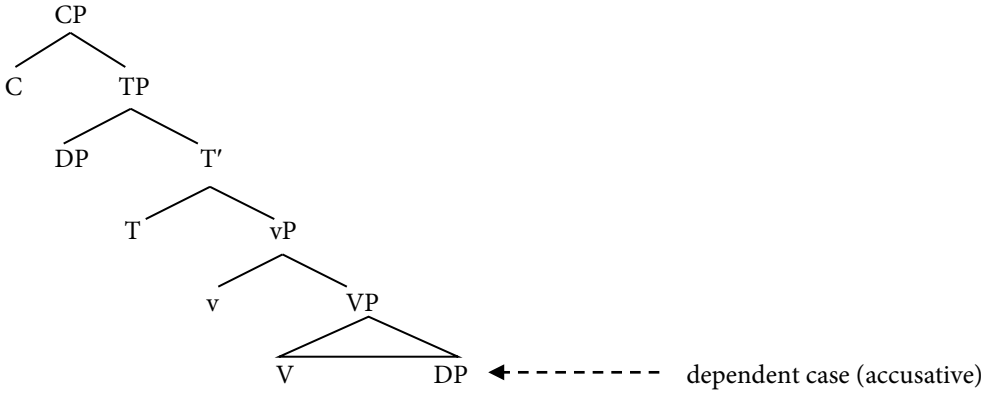
If soft domains are characterised as those domains in which unmarked case is not assigned, it follows that any DP within such a domain will remain active until we get to the next hard domain. At this point all unassigned cases must be assigned, including those remaining from the previous soft domain.

We start with a simple example: the Hungarian VP which is a soft domain spelled out when *v* merges into the structure. At this point, the object is eligible for unmarked case, being the sole DP in the domain. This case assignment is delayed and the DP remains active.

(19) Case assignment within VP (a soft domain):

We need to be more precise, however, about what case will be assigned to such still active DPs once we get to a hard domain. This will partly depend on whether the DP remains within the soft domain or whether it has moved. If the active DP remains in the soft domain, then it can interact with a DP in the higher hard domain and may therefore be eligible for dependent case assignment associated with the higher domain. This is what happens in the majority of nominative-accusative systems. In such languages, the VP is soft and so no unmarked case will be assigned to the object. When the universally hard TP is spelled out, the object is still active and therefore can be assigned the dependent case associated with TP, i.e. accusative.

(20) Case assignment within TP (a hard domain):



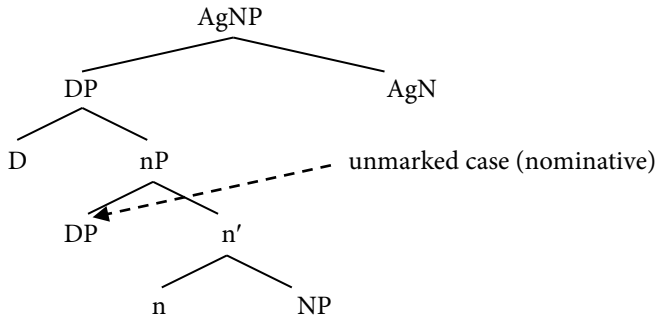
However, if there is no other DP in the hard domain, the still active DP will be assigned the unmarked case associated with the domain it is situated in. This is what happens in the case of Hungarian nominative possessors. As the lower nominal domain is soft and the higher one hard, we get the following sequence:

(21) Nominative case in possessive DPs:

a.	The possessor is merged into the structure along with the NP;	[$_{nP}$ DP n NP]]
b.	The determiner is merged with nP as its complement;	[$_{DP}$ D [$_{nP}$ n DP NP]]
c.	As D is a soft phase head, nP is spelled out, but case assignment delayed;	
d.	AgN merges with DP as its complement;	[$_{AgNP}$ [$_{DP}$ D [$_{nP}$ n DP NP]] AgN]
e.	As AgN is a hard phase head, DP is spelled out and all cases are assigned.	[$_{AgNP}$ [$_{DP}$ D [$_{nP}$ n DP$_{nom}$ NP]] AgN]

The possessor gets nominative case even though the assignment happens when the DP, which has dative as its unmarked case, is spelled out. This is because case assignment is relative to the position of the DP and the delayed unmarked case is not overridden by any case determined in the hard domain.

(22)



Compare this to what happens if the possessor is moved to the higher domain. Steps (21a), (b) and (c) will be the same, so we do not repeat them here:

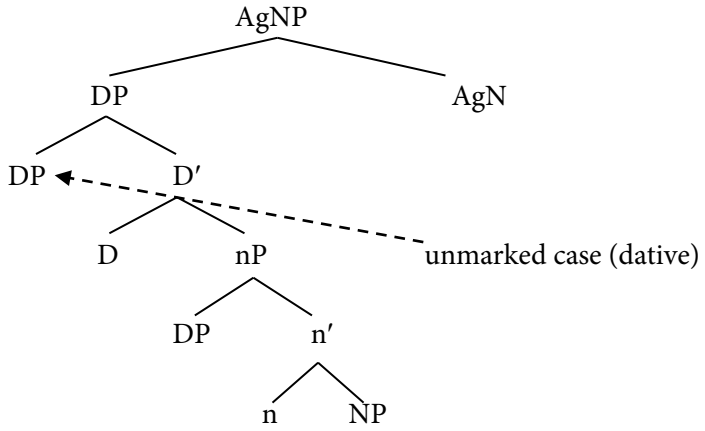
(23) Dative case in possessive DPs:

a. The possessor moves to the specifier of DP $[_{DP}$ DP D $[_{nP}$ DP n NP]]

b. AgN is merged and the DP is spelled out: all cases assigned $[_{AgnP}$ $[_{DP}$ DP$_{dat}$ D $[_{nP}$ DP n NP]] AgN]

Under these circumstances, the case assigned to the possessor is dative, the unmarked case associated with the Spell-Out domain of AgN, as the possessor moved into this domain.

(24)



Given that DP is a hard domain, once a possessor is assigned dative case, this cannot be changed even if it moves to a higher domain. Indeed, it can play no further role in the case system at all and so even if it is moved to a position higher than a subject, the subject will still be assigned unmarked nominative:

(25) *Jánosnak vágta ki Mari* $[_{AgNP}$ – *a fáját*].
John-DAT cut-PAST-3S down Mary-NOM the tree-3S-ACC
'It was John's tree that Mary cut down.'

In sum, 'case change' is the result of a delayed assignment of unmarked case. If, at the spell out of the next hard domain, conditions are such that another case can be assigned instead of the original unmarked one, that case will be assigned. If the conditions are not altered, the original unmarked case will be assigned. Case maintaining movement is merely the result of case assignment as this makes the recipient invisible to the case system.

8. A potential problem

The system we have outlined would appear to make the wrong predictions concerning case assignment to objects in Sakha. If we assume that the Sakha VP is a hard domain, as Baker does, as discussed following (3) above, then we predict that the unmarked case will be assigned to the object whether or not it remains inside the VP. Hence, the moved definite object will not be available for accusative case assignment. On the other hand, if we assume that the VP is soft, then the object will be eligible for accusative case assignment regardless of whether it moves. Hence the indefinite object should also be accusative.

However, Baker's account of definiteness-based differential case marking is problematic. Turkish displays similar definiteness-based differential object marking to Sakha, but in this case both the definite and indefinite objects remain inside the VP:

(26)

a. *Ahmet dün akşam pasta-yı ye-di.*
Ahmet yesterday evening cake-ACC eat-PAST
'Yesterday evening, Ahmet ate the cake.'

b. *Ahmet dün akşam pasta ye-di.*
Ahmet yesterday evening cake eat-PAST
'Yesterday evening, Ahmet ate cake.'

Kornfilt (2003)

Another analysis is therefore called for. We propose that definiteness-based differential object marking in Hungarian suggests a solution. In Hungarian, definite and indefinite objects are not assigned different cases, but they play a role in determining the agreement pattern on the verb. There is one agreement pattern associated with a definite object and another associated with intransitive verbs or those with indefinite objects. The pairing of indefinite objects with intransitive contexts suggests that the indefinite object appears to be invisible to the system and that only definite objects play a role in determining agreement marking. Applying this idea to Sakha and Turkish, we propose that indefinite objects are simply invisible in the case systems of these languages and hence are not eligible for case assignment. Thus, the movement of the definite object in Sakha is irrelevant to its case assignment. We claim that both Sakha and Turkish have soft VP domains and so the visible (definite) objects are assigned accusative, whether or not they move. This captures the facts more accurately than Baker's analysis and is fully compatible with our assumptions that unmarked case is not assigned in soft domains.

9. Conclusion

In this paper we have addressed the soft/hard domain distinction introduced in Baker (2015). This distinction is mechanically forced by specific properties of case for Baker, but why case behaves differently in different languages is not explained and the framework cannot be used to make predictions in this regard.

The main contribution of our paper is the proposal that unmarked case assignment in soft domains is delayed and all cases are assigned in hard domains. 'Case change' is just case assignment at another domain and case maintenance is simply the result of the already familiar case assignment mechanisms under movement. In order to present our account, we mainly used Hungarian data but our conclusions carry over straightforwardly to other languages. An apparent counterexample has been shown to involve independent factors.

Our analysis is not only more consistent with the principles of Dependent Case Theory, but the difference between case and movement also falls out from it. Recasting Marantz's (1991) Invisibility Principle in phase-theoretical terms has further increased the explanatory power of the theory.

References

Baker, M. 2015. *Case: its principles and parameters.* Cambridge: Cambridge University Press.

Chomsky, N. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language,* 1–52. Cambridge, Mass.: MIT Press.

Kornfilt, J. 2003. Scrambling, Subscrambling and Case in Turkish. In S. Karimi (ed.), *Word Order and Scrambling,* 125–155. Malden and Oxford: Blackwell.

Marantz, A. 1991. Case and Licensing. *ESCOL 91: Proceedings of the eighth Eastern States Conference on Linguistics*, 234–253. Columbus: Ohio State University.

Nakamura, H. 2002. Double subject, double nominative object and double accusative object constructions in Japanese and Korean. *Language, Information and Computation* 16: 358–369.

Newson. M and K. Szécsényi. 2020. Dative as the unmarked unmarked case in Hungarian. In M. Ziková, and M. Wągiel (eds.), *Linguistica Brunensia* 68(2): 27–44. Brno: Masaryk University.

Ryu, Byond-Rae. 2014 Semantic constraints on multiple case marking in Korean. In D. Gerland, C. Horn, A. Latrouite, and A. Ortmann (eds.) *Meaning and Grammar of Nouns and Verbs,* 77–112. Berlin, Boston: Düsseldorf university press. doi: 10.1515/9783110720075

Szabolcsi, A. 1983. The Possessor that Ran Away from Home. *The Linguistic Review* 3(1): 89–102.

Szabolcsi, A. 1994. The Noun Phrase. In F. Kiefer, and K. É. Kiss. (eds.) *The Syntactic Structure of Hungarian: Syntax and Semantics* 27: 179–127. New York: Academic Press.

Woolford, E. 2006. Lexical Case, Inherent Case, and Argument Structure. *Linguistic Inquiry* 37(1): 111–130. doi: 10.1162/002438906775321175.

LingBaW *Linguistics Beyond And Within 8 (2022), 167-179*

Do so replacement and the argument/adjunct distinction in Merge-based syntax*

Yuji Shuhama
Keiwa College, Shibata, Japan

Abstract
This study provides a novel look at *do so* replacement within the framework of the Bare Phrase Structure theory. Unlike the previous view of *do so* as a monolithic VP anaphor, I argue that *do so* is better analyzed as *do* and *so*, separately substituting for a functional Voice head and VP, respectively. This argument is supported by the observation of VP adverbs, the locative/directional interpretation of PPs, and the analysis of voice mismatch. The study consequently presents a more refined model of VP than the previous X-bar theoretic model, as it fulfills the structural requirement between complements and adjuncts.

Keywords: *do so*, Merge, locative/directional PPs, voice mismatch

1. Introduction

This study proposes a Merge-based account of *do so* replacement in English. *Do so* replacement was extensively explored in the era of X-bar theory, but it needs to be re-examined to conform to a more up-to-date theoretical context of the minimalist syntax, especially called Bare Phrase Structure (BPS) theory (Chomsky 1995, 2000). Among the issues in the BPS system, it seems theoretically controversial and thus worth discussing how adjuncts are adjoined to the binary structure and differentiated from complements in the Merge-based syntactic derivation.[1] Since *do so* replacement inevitably involves the issue of the structural distinction between complements and adjuncts, analyzing what is replaced by *do so* will lead

* I am grateful to two anonymous reviewers for their comments, which resulted in significant improvements of the manuscript. This research was funded by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 20K13146 for the author.

[1] In Fukui and Narita (2014), a number of important problems are listed for the future research of BPS theory. Adjunction is one such open issue, and various approaches based on pair-Merge, late-Merge, and antisymmetry are suggested for reconsideration whereby modifiers are allowed to be optionally adjoined to the modified constituents.

us not only to refine the structure of VP built by Merge, but also to uncover the essential nature of Merge in the BPS framework.

The paper is structured as follows: Section 2 briefly reviews the essential data of *do so* replacement and points out some structural shortcomings of the X-bar schema for the argument/adjunct distinction in VP. In Section 3, I demonstrate how Merge is driven by uninterpretable features and then propose that *do* and *so* independently substitute for a Voice head and VP, respectively. In Section 4, this proposal is tested by analyzing the voice mismatch between *do so* and its antecedent. Finally, Section 5 summarizes our findings and concludes the paper with an outline for future research.

2. What does *do so* replace?

2.1. Do so *replacing overt verbal phrases*

First, let us consider some generally accepted views of the relationship between *do so* and its verbal antecedent. A pivotal study by Hankamer and Sag (1976) discovered that *do so* requires an overt antecedent, while *do it* does not. The antecedent for *do so* and *do it* in (1), *stuff this ball through a 6-inch hoop*, appears overtly. In contrast, the antecedent of *do it* in (2a) can be pragmatically recoverable from the context, while that of *do so* in (2b) cannot. This difference in contextual recoverability clearly indicates that, unlike *do it*, *do so* must be strictly and syntactically controlled by the presence of its overt antecedent.

(1) A: I'm going to stuff this ball through a 6-inch hoop.
B: I don't believe that you can {do so / do it}.

Fu et al. (2001: 570)

(2)

a. [Hankamer attempts to stuff a 9-inch ball through a 6-inch hoop]
Sag: It's not clear that you'll be able to do it.

b. [Hankamer again attempting to pass 12" ball through 6" hoop]
Sag: #I don't think you can do so.

Hankamer and Sag (1976: 392, 418)

In addition to overtness, *do so* is known to require its antecedent to be larger than a single V. The contrast between (3a, b) and (3c) shows that *did so* can substitute for verbal phrases *gave a speech* and *gave a speech on the 30th of June*, but not for the past verb *gave* itself. This leads us to a generalization that when the antecedent is targeted for *do so* replacement, it should be a verbal constituent including at least a verb's complement like *gave* plus *a speech* in (3).

(3) John gave a speech on the 30th of June in Madrid, and…

a. Mary did so in Valencia.

b. Mary did so on the 27th of September in Valencia.

c. *Mary did so a statement later that day.

Poole (2011: 41)

The next example (4) may appear to be a counterexample to the claim that a verb itself cannot be replaceable. However, if we follow Boeckx and Stjepanović (2001), who argue for the PF analysis of head movement on the basis of pseudogapping constructions like (4), *did the cookies* should be better treated as a remnant object moving out of the VP-ellipsis site rather than *did* directly replacing *ate*.[2]

(4) Debbie ate the chocolate, and Kazuko did the cookies.

Boeckx and Stjepanović (2001: 346)

The apparent contrast between ungrammatical *did so* in (3c) and grammatical *did* in (4) is not due to the different anaphoric nature of *do so* and *do*, but supposedly due to the underlying mechanism of *do so* replacement and VP-ellipsis. This argument leads us to the following question: As distinct from VP-ellipsis, what makes *do so* unable to target a single V for replacement? Although it is beyond the scope of this study to discuss this in detail, I will leave it as an important open question for future investigation.

2.2. X-bar theoretic account

Do so can substitute for a verbal constituent, but whether the antecedent can be replaced by *do so* or not seems to depend on the property of verbs within the targeted constituent. For example, even though both (5a) and (5b) have the same V-DP-PP sequence in the first conjunct (i.e., *read a book in the attic* and *put a book on the table*), *did so* in (5a) can replace a sequence of V plus DP, *read a book*, while *did so* in (5b) cannot substitute for the categorically identical sequence of V plus DP, *put a book*. As (5c) shows, (5b) becomes grammatical if *did so* refers to a broader sequence of its antecedent including a PP, *put a book on the table*. To account for the contrast regarding the types of preceding verb, we need to structurally distinguish two positions for PPs: one for adjunct PPs, which can be exempt from *do so* replacement, and the other for complement PPs, which are necessarily part of the replaced constituent.

(5)

a. Mary read a book in the attic, and John did so in the garden.
b. *Mary put a book on the table, and John did so on the floor.
c. Mary put a book on the table, and John did so, too.

---

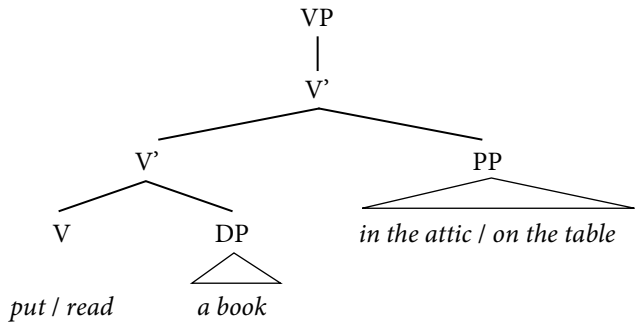
[2] Based on Lasnik's (1999) analysis, Boeckx and Stjepanović point out that Object Shift must apply and then the remnant VP is elided in pseudogapping constructions, while both Object Shift and the head movement of V apply necessarily in the standard VO order. Each of the examples with bracketed derivational steps is shown in (i-a) and (b), respectively.

(i) Debbie ate the chocolate, and …
a. Kazuko did [the cookies ~~[$_{VP}$ ate *t*]~~] (=(4))
b. Kazuko [drank$_V$ [milk$_i$ [$_{VP}$ t_V t_i]]]

They conclude that after Object Shift applies in overt syntax, VP-ellipsis and V-raising apply as PF operations, which compete with each other (i.e., either of them selectively takes place, not both) to yield either pseudogapping or VO-ordered sentences.

One solution proposed in X-bar theory was to replace a V-bar (V') with *do so* under the assumption that VPs are uniformly analyzed to have a structure similar to (6).

(6)



The crucial problem of assuming the X-bar structure above is that this structure does not provide any reason why (5a) and (5c) are grammatical but (5b) must be ruled out. Moreover, the current BPS theory presumes that it is desirable to eliminate the X-bar component of Universal Grammar (UG) and reduce structure building to the operation Merge.[3] The X-bar approach assumes that adjuncts in English are base-generated adjacent to the right of V' or VP.

Another problem is, as previously argued by Larson (1988) and Pesetsky (1995), the approach outlined above cannot account for essential c-command relations between the adjuncts and the complements of the verb. Typical diagnostics include negative polarity item (NPI) licensing and quantifier-bound pronouns in (7). For example, (7a) shows the licensing of an NPI *any* by the preceding negative element *no* in a higher position. Each of the tested examples in (7) indicates that adjuncts are c-commanded by the postverbal complements.[4]

(7)

a.	John saw *no* student in *any* classroom.	Negative polarity items
b.	John visited *everyone*$_i$ on *his*$_i$ birthday.	Quantifier-bound pronouns

The c-command relation between the arguments and adjuncts holds in the structure derived via obligatory movements to satisfy featural requirements rather than optional movements

---

[3] For example, Chomsky (1995: 209) proposes that syntactic computations should conform to the Inclusiveness Condition, which states that no new features can be added in the course of the derivation and only the items from the numeration can be used. Bar-levels as well as X^0- and XP-levels are merely theory-internal distinctions within X-bar theory. Since they cannot be construed as intrinsic features of any items from the numeration, such intermediate levels are considered to be abandoned for the structure-building in the BPS framework.

[4] You may wonder if quantifier raising enables the quantified objects to c-command adjuncts with licensed/bound expressions even in the X-bar structure of VP like (6). This option, however, does not seem plausible. Quantified phrases such as *no linguist*/*conference room* covertly adjoin to TP, and if they licensed *any conference room*/*linguist* by c-commanding the domain containing them at LF, both (ii-a) and (b) would be grammatical. Since (ii-b) is eventually ruled out, I do not adopt the c-command relation at LF to account for the structural distinction between arguments and adjuncts.

(ii)

a. John spoke to Mary about *no linguist* in *any conference room*.

b. *John spoke to Mary about *any linguist* in *no conference room*. (Cinque 2006: 141)

afterwards such as stylistic fronting. This is suggested by the contrast between the acceptable and ungrammatical results of pronominal binding in (8) and (9). In (8), the embedded subject *two men* is ECMed due to *v*'s EPP feature and φ-valuation, and behaves like an object of *prove* to bind *each other* in the matrix adjunct.

(8)

a. The DA proved two men to have been at the scene of the crime during each other's trial.

b. *The DA proved there to have been two men at the scene of the crime during each other's trial.

Lasnik (1999)

(9) shows the case of optional movements. Unlike (9a), the fronted PP *near Dan* in (9b) cannot refer to *he* from its landing site even though it appears to be controlling its referent within the root sentence.

(9)

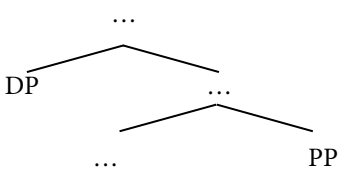
a. Near him$_i$, Dan$_i$ saw a snake t.

b. *Near Dan$_i$, he$_i$ saw a snake t.

Reinhart (1981)

Since the syntactic relations between complements and adjuncts cannot be properly analyzed or guaranteed in the X'-based VP structure in (6), an alternative VP structure must be considered along the lines of Larson's (1988) proposal, in which adjuncts are embedded more deeply below the complements of verbs. In other words, putting aside some details of categorial information, (10) can be a more promising VP format than (6) with respect to the c-command relation between higher complements and lower adjuncts.

(10)



3. Merge-based analysis of *do so*

3.1. Merge *as a feature-matching operation*

In order to identify the internal structure of VP fulfilling the structural requirement observed so far, let us begin by closely observing the transitive location verbs such as *put* and *smear*. These verbs are known to be three-place predicates: *put* and *smear* take one external AGENT, and two internal THEME and LOCATION arguments. As required by such argument specification exemplified in (11), a sentence (or a clause) whose core predicate is *put* or *smear* must have all the three arguments present. Otherwise, lacking any one of required arguments makes the sentence/clause ungrammatical, as shown in (12).

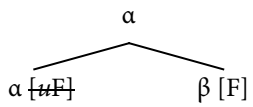
(11) *put*, *smear*, etc.: <1, 2, 3> (AGENT (THEME (LOCATION)))

(12)

a. *There put a book on the table. <$ø_{AG}$, 2_{TH}, 3_{LOC}>
b. *Mary put on the table. <1_{AG}, $ø_{TH}$, 3_{LOC}>
c. *Mary put a book. <1_{AG}, 2_{TH}, $ø_{LOC}$>

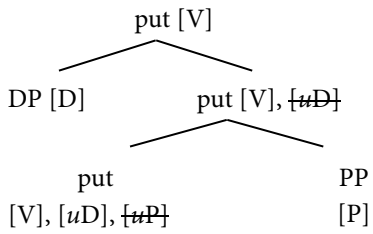
Then, let us consider how Merge applies to form a VP with all V's required arguments fulfilled. Merge occurs when there is a syntactic object (SO) carrying an uninterpretable feature [*u*F] that needs to be satisfied. Suppose that there are an SO α carrying [*u*F] and another SO β carrying a relevant feature [F]. By merging two SOs α and β into a new SO γ, the α's [*u*F] matches the β's [F] and then it becomes deleted, otherwise the undeleted [*u*F] causes the derivation to crash. The feature-matching model of Merge described so far is graphically summarized in (13), where α projects for a new SO as a result of feature-matching.

(13)



It follows from the above view of Merge as a feature-matching operation that if there is more than one uninterpretable feature, which is actually the driving force of Merge, Merge must apply more than once. Since *put* and *smear* require to take two internal arguments (THEME and LOCATION), suppose that they carry uninterpretable D- and P-features (coded as [*u*D] and [*u*P]). (14) demonstrates how Merge proceeds step-by-step to form a VP like *put a book on the table*: the first Merge is driven by the [*u*P] of *put* to match the [P] of a PP *on the table*, and then the second Merge takes place in turn to match the unchecked [*u*D] of *put* and the [D] of a DP *a book*.

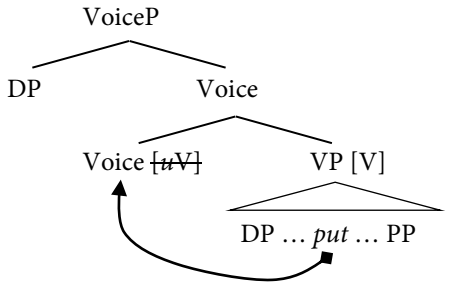
(14)



3.2. Do so *and* VoiceP

Now that all the internal arguments are fulfilled by Merge, the external AGENT argument is then introduced to the structure by merging VP and a Voice head, which is responsible for specifying the voice for a state/event as active, passive, or middle. Suppose the following three properties of Voice: (i) Voice has an uninterpretable V-feature [*u*V] to be matched/deleted, (ii) a Voice head is affixal, and (iii) the Voice in (15) is valued as active. Since an affixal head itself cannot stand independently without being supported by some other independent morphemes, verbs such as *put* and *smear* raise from VP to Voice to be a verbal host of the Voice affix. As a result, we have obtained all the necessary external and internal arguments in the right order: there are an AGENT DP and the verb within VoiceP, and the THEME DP and LOCATION PP within VP.

(15)



Then, what structure, or which constituent, does *do so* substitute for within the VoiceP in (15) above? Instead of dealing with *do so* as a monolithic verbal proform, it seems to be structurally well-grounded to analyze *do* and *so* separately for the following two observable reasons: (i) the placement of VP-adverbs, and (ii) *do so* can be characterized by its own voice.

If *do so* is better analyzed as a separated realization of a higher Voice head and a lower VP instead of having a rigid, monolithic structure, it is theoretically predicted that VP-adverbs can intervene between *do* and *so* to modify an event substituted by *do so*. This prediction is quite consistent with the following samples of VP-adverbs obtained from the web: (16) shows that *do so* can be intervened by VP-adverbs such as *amazingly*, *gracefully*, and *reluctantly*.[5]

(16)

a. The way it handles open-wide exploration – which it does *amazingly* so – has blinded many people to its faults I think.
b. Now my Lady I humbly suggest you take a look at the video I posted, here is a woman that lets herself expose to the 3D, and in my opinion does *gracefully* so, maintaining her energy of unconditional love.
c. "Okay," Patton says, urging Virgil to let go of him and he does *reluctantly* so.

(The data obtained from the web; italics added)

Turning now to voice characterization, in (17), for example, *do so* appears in active voice (*done so* and *doing so* substituting for *contacted Mary* and *discussing the issue*, respectively), even though the antecedent is in passive voice (*Mary was discussed* and *the issue had been discussed*).

(17)

a. Mary was contacted by the same man in Boston who had *done so* in New York.
b. Because the issue had been discussed so thoroughly in our committee that afternoon, we were asked not to waste time *doing so* again that night.

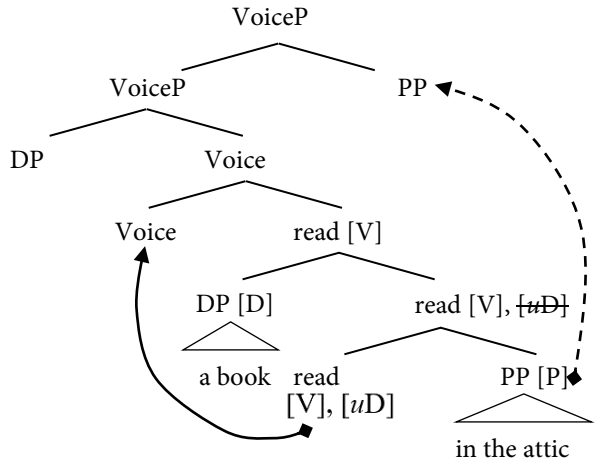
Bouton (1969, cited in Bruening 2018)

Here, another question arises as to the structural analysis of other VPs: how is a VP formed if it is headed by other types of transitive verbs like *read* and *write*? These verbs are two-place predicates (i.e., predicates that require both AGENT and THEME arguments), so unlike locative transitive verbs like *put* and *smear*, they do not have to take locative PPs (e.g., *in the*

[5] I would like to thank Prof. Marcel den Dikken for bringing the position of VP-adverbs to my attention. Each sample in (16) is cited from several sites, accessed on May 30, 2022, whose URLs are listed as follows:
(a) https://forums.escapistmagazine.com/threads/why-is-zelda-botw-considered-so-amazing.142156/
(b) https://aishanorth.wordpress.com/2013/09/10/the-manuscript-of-survival-part-352/
(c) https://www.wattpad.com/520999871-spicy-sanders-sides-one-shots-do-you-think-i-look

attic) to satisfy their argument specification. (18) is the VoiceP structure I would like to propose in order to analyze the verbal constituent *read a book in the attic*.

(18)



Although Merge is considered to apply freely to combine any two SOs (Chomsky 2004, 2015), in (18) *read* first takes the adjunct PP, and then it successively takes the required THEME argument to match and delete its uninterpretable D-feature. By doing so, a proper c-command relation is guaranteed between the higher complement DP and lower adjunct PP within VP. When the derivation proceeds and Voice merges VP to form VoiceP, a head movement of *read* takes place to host an affixal Voice head, as mentioned in (15). Since Voice in the above case is valued active, an AGENT DP is then externally merged to project VoiceP.[6]

I would also like to assume an optional movement of the VP-internal adjunct PP to the higher VoiceP-domain (signified as a dotted arrow in (18)). This movement allows *do so* replacement to target verbal antecedents in different sizes. For example, when the PP *in the attic* moves out of VP to adjoin to VoiceP, the target of *do so* can be narrowed down from the full VP *read a book in the attic* to its subpart *read a book* only, since the moved PP becomes invisible for *do so* to search for as its target within VP-domain. This evacuation from VP to the higher domain, however, cannot be applied to a complement PP (e.g., *put a book* [PP *on the table*]). Since VP is transferred at the completion of a *v*- phase, I assume that the structural relations to check V's uninterpretable features within VP are preserved and transmitted to the phonological/semantic interfaces for further computation.[7]

---

[6] A reviewer pointed out the *do so* replacement including temporal adverbials such as *in the morning* such as *Mary will cook the potatoes in the morning, and Susan will do so, too*. The *do so* data presented throughout this paper are limited to event-related because I believe the complement/adjunct distinction in Neo-Davidsonian event semantics to be consistent with my structural analysis of complement/adjunct PPs. The above issue of course will be explored in the future research as well as how event modification by tense and temporal adverbials is framed in Neo-Davidsonian semantics.

[7] A reviewer wonders why a complement PP does not merge VoiceP optionally. An anaphoric contrast in (iii) shows that complement PPs fall within the same binding domain as objects while adjunct PPs do not.

(iii)

a. John put Bill$_i$ in front of a picture of himself$_i$.

b. ??John met Larry$_i$ in front of a statue of himself$_i$. (Hestvik 1991: 463, 465)

I assume for the time being that the former are kept connected to the VP-domain due to the [*u*P] feature of V.

3.3. Two positions for PPs

This optional-movement analysis of adjunct PPs to VoiceP is supported by and closely related to some recent views of the interpretation of PPs. One of the latest studies of PPs in line with our purpose here is Milway (2015), who analyzed Locative and Directional interpretation of PPs in multiple VP-layers. For example, (19) is interpreted ambiguously: in the locative interpretation, the PP *behind the screen* means that Jamie was behind the screen, and he ran there. In the directional interpretation, on the other hand, the same PP implies that at first Jamie was somewhere else, and then he ran towards back of the screen.

(19) Jamie ran behind the screen. Locative/Directional

Milway (2015: 10)

However, when *behind the screen* is fronted as in (20), the interpretation is fixed to locative, not directional. This fact suggests that the interpretation of PPs can be affected and restricted by structural height of PPs.

(20) Behind the building, Andrew dances
… but in front of it, he stands perfectly still. Locative/*Directional

Milway (2015: 12)

Then, let us observe how PPs are interpreted when they modify *do so*. In (21), the PP *on the track* modifies *do so*, and it is interpreted as locative, rather than directional. *Did so on the track* can actually be interpreted ambiguously as either (a) or (b), but in both interpretations, the PP *on the track* modifying *do so* is interpreted as locative. In addition to PP-fronting in (20), this locative-oriented interpretation of PPs also indicates that locative PPs are structurally higher than directional PPs. Based on Milway's proposal summarized in (22), I assume that the PPs exempt from *do so* replacement are adjoined to VoiceP, while the PPs replaced by *do so* are within VP.[8]

(21) The soccer players ran between the nets and the sprinters did so on the track.
a. … ran [$_{loc}$ between the nets] and … ran [$_{loc}$ on the track]
b. … ran [$_{dir}$ between the nets] and … ran [$_{dir}$ between the nets] [$_{loc}$ on the track]

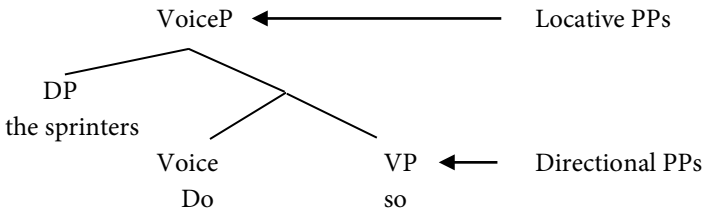
Milway (2015: 14)

[8] As cited in Cinque (2010: 15–16, fn. 12), Stringer's (2006) examples of locative PPs moving out of the VP domain indicate that locative PPs consist of a layered structure with higher functional PATH and lower lexical PLACE heads. Based on the layered PP structure exemplified in (i), Stringer points out that a directional reading of locative PPs such as *on the pitch* in (ii) becomes impossible when the PPs are moved to a higher, focused position, as shown in (iii).

(i) jump [$_{PathPP}$ from [$_{PlacePP}$ in [$_{LocNP}$ front [$_{PP}$ of [$_{DP}$ the train]]]]]
(ii) Zidane ran *on the pitch*. Locative/Directional
(iii) It was *on the pitch* that Zidane ran. Locative/*Directional

The lack of directional reading in (iii) is due to a covert PathP being unable to be licensed under the surface adjacency to the verb (*ran*, in this case).

(22)



Milway (2015: 13)

My proposal is summarized in (23). Since *do so* can have its own voice different from the voice of the antecedent, I analyze *do* independently as a pro-Voice head, and *so* as a pro-VP. Once the internal arguments required by verbs and the adjuncts modifying them are merged to maintain the c-command relation within VP, the adjuncts can internally merge VoiceP, just like *in the attic* in (23b). By doing so, PPs can escape from the target VP domain of *so*-substitution, and the PPs modifying *do so* from VoiceP can get interpreted as locative.

(23)

a. [$_{VoiceP}$ Mary [Voice+ put [$_{VP}$ a book [~~put~~ on the table]]]]
| do | so |

b. [$_{VoiceP}$ Mary [Voice+read [$_{VP}$ a book [~~read in the attic~~]]] in the attic]
| do | so |

4. *Do so* and Voice Mismatch

Lastly, let us consider another case of voice mismatch between *do so* and its antecedent to see if my analysis can account for it. The examples (24) show the contrast of active-middle alternation: the antecedent of *do so* in (a), *melted the glass*, seems to be replaceable with the middle voice sentence *it would do so*, which stands for "the glass would melt," while in (b) such active-middle alternation is not possible between the active VP *killed Mary* and the middle counterpart *she did so*, which is intended to mean "Mary died". (24a) used to be accounted for by assuming the lexical semantic structure of causation like (24), CAUSE X TO DO SOMETHING. However, it was proven to be problematic because it could not explain why only the middle interpretation of (*she*) *did so* in (24b) is ruled out if *melt* and *kill* can be equally characterized with the same semantic structure as in (25).

(24)

a. Floyd melted the glass though it surprised me that {he / it} would do so.
= Floyd would melt the glass / the glass would melt

b. John killed Mary, and it surprised me that {he /*she} did so.
= John killed Mary / *Mary died

Foder (1970: 429), Dowty (1979: 240)

(25) melt / kill : CAUSE X to melt / die

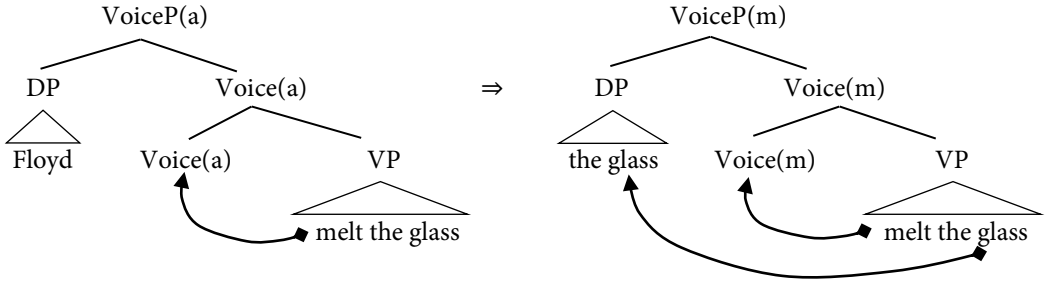
Alternatively, my analysis starts with identifying the argument specification of verbs. Take *melt* in (24a), for example. *Melt* is an unaccusative verb and, as specified in (26a), it only needs one internal THEME argument. The derivation proceeds as (26b) shows: *melt* Merges a

DP *the glass* in order to get its uninterpretable D-feature [*u*D] checked within VP. Then, a Voice head Merges the VP. Since this Voice head is valued as active, it takes *Floyd* as the subject by external Merge, and then the VoiceP projects. When *do so* replacement applies, *so* refers to the anteceding VP *melt the glass*. Since the Voice head for *do so* can be valued independently of the Voice value for the antecedent, in the right structure it is valued as middle, in other words, non-causative. The middle Voice head triggers the internal Merge of *the glass* from the pro-VP domain to the subject position in VoiceP. As a result, we have *the glass* in VoiceP, a middle Voice head incorporated with *melt*, and a VP below them, each of which corresponds to *it*, *do*, and *so*, respectively.

(26)

a. melt: <2> (THEME)

b.



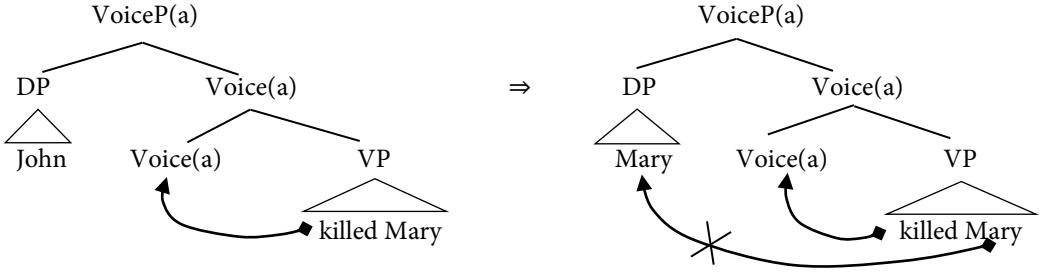
'Floyd melted the glass …' '… that it would do so.'

On the other hand, the same does not apply to another example, *John killed Mary* in (24b). As (27a) shows, *kill* needs one external AGENT argument and another internal PATIENT argument. Since *kill* needs an AGENT argument as its subject, the Voice must be valued as active to have the agentive subject by external Merge. When *do so* replacement is applied, the antecedent VP, *kill Mary*, must be referred to by *so*, but the Voice for *do so* should be active in order for the external AGENT argument of *kill* to be fulfilled by external Merge, not by internal Merge of *Mary* from inside the embedded VP.

(27)

a. kill: <1, 2> (AGENT (PATIENT))

b.



'John killed Mary, …' '… that *she did so.'

5. Conclusion

In this paper, I have argued that unlike the previous treatment of *do so* as a monolithic VP anaphor, *do so* is better analyzed as a structural compound of separate *do* and *so*. Let us review some main points to summarize the above discussion. As for the transition from the X-bar to

BPS theory, the first problem I pointed out was that the c-command relation between the complements and adjuncts is not maintained in the X'-based VP model. Then, in order to overcome this structural inadequacy, I proposed the VoiceP structure embedding a VP based on the observation of VP-adverbs and voice characterization of *do so* and its antecedent. *Do* is a morpo-phonological realization of Voice head, and *so* substitutes for VP. Adjuncts in VP can be internally Merged up to VoiceP to become exempt from the targeted domain of *do so* replacement and get interpreted as locative. Finally, two cases of voice mismatch with *melt* and *kill* were examined to test my proposal.

Important tasks to be explored for future research include comparing the underlying syntax of *do so* replacement and VP-ellipsis. As mentioned in Section 2.1, *do so* differs from pseudogapping in that it basically cannot target a single V for replacement. However, there are in fact several examples where *do so* appears to substitute for V, including Milway's (2015) example (21a) observed in Section 3.3. Another topic is to see to what extent my proposal of *do so* replacement can be similarly applied to other kinds of proforms such as *one* substitution. Investigating these further issues will contribute to a more extensive, clarified picture of VP/NP structures and related anaphoric phenomena.

References

Boeckx, C., and S. Stjepanović. 2001. Head-*ing* towards PF. *Linguistic Inquiry* 32(2): 345–355.

Bouton, L. 1969. Identity constraints on the *do-so* rule. *Research on Language and Social Interaction* 1: 231–247.

Bruening, B. 2018. Passive *do so*. *Natural Language and Linguistic Theory* 37: 1–49.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. 2000. Minimalist inquiries: The framework. In R. Martin, D. Michaels, and J. Uriagereka (eds.), *Step by step: Essays in minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, N. 2004. Beyond Explanatory Adequacy. In A. Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, Vol. 3, 104–131. Oxford: Oxford University Press.

Chomsky, N. 2015. Problems of Projection: Extensions. In E. Di Domenico, C. Hamann, and S. Matteini (eds.), *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*, 3–16. Amsterdam: John Benjamins.

Cinque, G. 2006. *Restructuring and Functional Heads: The Cartography of Syntactic Structures*, Vol. 4. New York: Oxford University Press.

Cinque, G. 2010. Mapping Spatial PPs: an introduction. In G. Cinque, and L. Rizzi (eds.), *Mapping Spatial PPs. The Cartography of Syntactic Structures*, vol. 6, 3–25. Oxford, Oxford University Press.

Dowty, D. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Foder, J. 1970. Three Reasons for Not Deriving "Kill" from "Cause to Die". *Linguistic Inquiry* 1(4): 429–438.

Fu, J., Roeper, T., and H. Borer. 2001. The VP within Process Nominals: Evidence from Adverbs and the VP anaphor *do-so*. *Natural Language & Linguistic Theory* 19(3): 549–582.

Fukui, N., and H. Narita. 2014. Merge, Labeling, and Projection. In A. Carnie, Y. Sato, and D. Siddiqi (eds.), *The Routledge Handbook of Syntax*, 3–23. London and New York: Routledge.

Hankamer, J., and I. Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 7(3): 391–426.

Hestvik, Arild. 1991. Subjectless Binding Domains. *Natural Language and Linguistic Theory* 9: 455–496.

Larson, R. 1988. On the Double Object Construction. *Linguistic Inquiry* 19(3): 335–391.

Lasnik, H. 1999. Chains of Arguments. In S. D. Epstein, and N. Hornstein (eds.), *Working Minimalism*, 189–215. Cambridge, MA: MIT Press.

Milway, D. 2015. Directionalized locatives: a label theoretical account. Ms., University of Toronto.

Pesetsky, D. 1995. *Zero Syntax*. Cambridge, MA: MIT Press.

Poole, G. 2011. *Syntactic Theory*. 2nd edition. London: Red Globe Press.

Reinhart, T. 1981. Definite NP-Anaphora and C-Command Domains. *Linguistic Inquiry* 12: 605–635.

Stringer, D. 2006. Typological Tendencies and Universal Grammar in the Acquisition of Adpositions. In P. Saint-Dizier (ed.), *Syntax and Semantics of Prepositions*, 57–68. Dordrecht: Springer.

LingBaW *Linguistics Beyond And Within 8 (2022), 180-199*

On bare and non-bare temporal names in Romanian*

Mihaela Tănase-Dogaru
University of Bucharest, Romania

Abstract
The present paper investigates punctual vs. habitual readings of Romanian proper temporal names of the type *luni* 'Monday' vs. *lunea* 'Monday.DEF'. These readings are associated with the absence vs. presence of the definite article (Franco and Lorusso 2022). The paper makes two major claims. Firstly, following Longobardi (1994, 2005), and Franco and Lorusso (2020), the paper claims that with bare, i.e., definiteless, proper time names, N-to-D movement triggers individual-like reference, which, in turn, explains why the event is interpreted as punctual. Secondly, the paper shows that the structure of proper temporal names is complex, in the sense that it contains the classifier *zi* 'day', thus paralleling the structure of complex descriptive proper names of the type 'the planet Venus' (see van Riemsdijk 1998, Cornilescu 2007 a.o.). This classifier is shown to be overt when there is no N-raising, and silent when N raises to D in the structure of proper temporal names.

Keywords: temporal names, classifiers, N-raising

1. Introduction and aims

The aim of the present paper is two-fold. On the one hand, the paper aims at investigating punctual readings (1) versus habitual readings (2) of proper temporal names in Romanian, focusing on names of days (see Franco and Lorusso 2020 for Italian): [1]

(1) *Merg* *la film* *marți.*
go.PRES.1SG to movie Tuesday
'I'm going to the movies on Tuesday.' = one specific Tuesday

* I gratefully acknowledge the comments and suggestions of two anonymous reviewers, which helped me to write a much-improved version of the paper. I thank Anna Bondaruk, Marcel den Dikken, and Jacek Witkoś, as well as the audience at LingBaW 2021 for their very helpful comments and suggestions. All remaining errors are my own.

[1] The abbreviations used in the paper are the following: ACC – accusative; CL – clitic; DAT – dative; DEF – definite article; DOM – differential object marking; FEM – feminine; FUT – future; IMPERF – imperfective; MASC – masculine; PL – plural; PRES – present; REFL – reflexive; SG – singular

(2) *Merg* *la film* *marțea.*
go.PRES.1SG to movie Tuesday.DEF
'I go to the movies on Tuesdays.' = every Tuesday[2]

As shown in example (2), when names of days appear as adverbials modifying an event, the event is interpreted as a *habitual* one if the temporal name is accompanied by the definite determiner. The example in (1), on the other hand, shows that the event is interpreted as a *punctual* one if the temporal name is bare.

Following Longobardi (1994, 2005), and Franco and Lorusso (2020), the paper claims that with bare proper time names, N-to-D movement triggers individual-like reference, which, in turn, explains why the event in interpreted as punctual. N-to-D movement is apparent in the syntax of (3 a, b). (3a) illustrates the base position of the noun *marți* / 'Tuesday', viz. the N position. In (3b), the noun *marți* / 'Tuesday' raises across the adjective *viitoare* / 'next' and incorporates the definite article.

(3)
a. *Merg* *la film* *viitoarea* **(zi de)* *marți.*
go.PRES.1SG to movie next.DEF *(day of) Tuesday
'I'm going to the movies next Tuesday.'
b. *Merg* *la film* *marțea* *viitoare.*
go.PRES.1SG to movie Tuesday.DEF next
'I'm going to the movies next Tuesday.'

The process of N-raising in (3b) is accompanied by an interesting phenomenon, the obligatory emergence of the classifier *zi* 'day' (to be discussed in section 3.2). Therefore, the second aim of the paper is to show that, with names of days, Romanian uses *zi* 'day' as a

[2] An anonymous reviewer points out that the definite, i.e. *marțea* 'Tuesday.DEF' should not be interpreted as universal, but as generic. Similarly, indefinite temporal nominals like *marți* 'Tuesday' can also have a generic interpretation. However, according to my and my informants' judgments, the most natural interpretation of the pair (1)-(2), in an out-of-the-blue scenario, is punctuality (one event) versus habituality (a sequence of events).

The anonymous reviewer suggests that the right generalization with respect to the data in (1) and (2) seems to be that while the definite forms get interpreted as habitual, bare temporal nominals are the elsewhere realization, in the sense that they can be both habitual and punctual. They offer the example in (a) as an illustration of the fact that bare proper temporal names can have a habitual interpretation (I have taken out the glosses for simplification):

(i) *Mariei îi place rutina. Are un program săptămânal pe care îl respectă cu strictețe.* ***Luni*** *se duce la teatru,* ***marți*** *se duce la operă,* ***miercuri*** *are ore de balet...*
'Maria likes routine. She has a weekly schedule that she strictly sticks to. On Monday she goes to the theatre, on Tuesday she goes to the cinema, on Wednesday she takes ballet classes…'

However, my and my informants' judgments point to the fact that even when the bare temporal name is embedded in a larger scenario, i.e. as opposed to out-of-the-blue sentences, the habitual interpretation is likely to pair with the definite proper temporal name:

(ii) *Mariei îi place rutina. Are un program săptămânal pe care îl respectă cu strictețe.* ***Lunea*** *se duce la teatru,* ***marțea*** *se duce la operă,* ***miercurea*** *are ore de balet...*
'Maria likes routine. She has a weekly schedule that she strictly sticks to. On Monday.DEF she goes to the theatre, on Tuesday.DEF she goes to the cinema, on Wednesday.DEF she takes ballet classes…'

classifier (see also the analysis of restrictive appositives in van Riemsdijk 1998 and complex or descriptive proper names in Cornilescu 2007). The paper shows that the classifier is overt when there is no N-raising (5); it is silent, however, when N raises to D (6).

(4) Merg la film viitoarea *(zi de) marți.
go.PRES.1SG to movie next.DEF (day of) Tuesday
'I'm going to the movies next Tuesday.'

(5) minunata *(zi de) marți
wonderful.DEF day of Tuesday
'the wonderful Tuesday'

(6) marțea minunată
Tuesday.DEF wonderful
'the wonderful Tuesday'

It will be shown that proper temporal names take classifiers, as illustrated in (4), while the role of 'de'/'of' is that of a partitive marker, assigning abstract genitive case (see Tănase-Dogaru 2008, 2009, 2017).

2. The data

In Romanian, the set of proper temporal adverbials that modify verbal events differ in interpretation. This set includes days of the week (e.g., *luni*, Monday, *marți*, Tuesday, etc.).

(7) *Merg la magazine lunea.*
go.PRES.1SG at shop Monday.DEF
'I go shopping on Mondays.'

(8) *Merg la magazine luni.*
go.PRES.1SG at shop Monday
'I'm going shopping on Monday.'

When the temporal modifier is accompanied by the definite determiner (7), the event is interpreted as a habitual one, while when the proper temporal modifier lacks a definite determiner and it, therefore, appears 'bare' (8), the event is perceived as punctual, non-iterable[3].

[3] Not all native speakers share the judgements concerning the data in (7) and (8). However, in out-of-the-blue sentences or natural question-and-answer pairs of the type in (a) (where speaker A inquires about speaker B's arrangement on a specific Monday, i.e. next Monday).

(iii) A: *Ce faci luni?*
what do.PRES.1SG Monday?
'What are you doing on Monday? / this Monday?'
B: *Merg la film luni / Luni merg la film.*
go.PRES.1SG at movie Monday / Monday go.PRES.SG at movie
'I'm going to the movies on Monday.'

The same variation can be found in Italian (Franco and Lorusso 2022). (9b) shows the habitual reading, encapsulated by the definite proper temporal name, while (9c) shows the punctual reading, with the bare temporal name. Unlike Romanian, Italian also has the variant in (9a) for the habitual reading, featuring the genitive-assigning preposition *di* (Franco and Lorusso 2022: 2):

(9)

a. *vado al cinema di lunedì* = habitual reading
go.PRES.1SG to.the cinema of Monday
'I (usually) go to the cinema on Mondays.'

b. *vado al cinema il lunedì* = habitual reading
go.PRES.1SG to.the cinema DEF Monday
'I (usually) go to the cinema on Mondays.'

c. *vado al cinema lunedì* = punctual reading
go.PRES.1SG to.the cinema Monday
'I go to the cinema next Monday.' Franco and Lorusso (2022: 2)

Unlike Italian, where the contrast between the punctual and the habitual reading holds only for irrealis contexts (see Franco and Lorusso 2022), in Romanian the contrast holds for other tenses, the past (10) or the future (11), which indicates that it is the peculiar behavior of proper temporal adverbs that triggers the asymmetry in the aspectual interpretation of the event, and not the interpretation of the tense involved, i.e., realis or irrealis.

(10) *Mergeam la magazin lunea / luni.*
go.IMPERF.1SG at shop Monday.DEF / Monday
'I went shopping on Mondays / I went shopping on Monday.'

(11) *Voi merge la magazin lunea / luni*
will.1SG go to shop Monday.DEF / Monday
'I'll go shopping on Mondays / I'm going shopping on Monday.'

Another relevant observation concerns time adverbs. Time adverbs like *ieri* 'yesterday' or *azi* 'today' cannot be employed with definite determiners, irrespective of their realis or irrealis orientation (12)-(13). They can convey only a punctual interpretation of the event, just like the 'bare' adverbials in (8) (for the Italian data see Franco and Lorusso 2022:6).

(12) *Am mers la magazin ieri / *ieriul.*
have.1SG gone to shop yesterday / yesterday.DEF
'I went shopping yesterday.'

In contrast, in question-and-answer pairs of the type in (b), the habitual interpretation is triggered (speaker A inquires about what speaker B usually does)

(iv) A: *Ce faci lunea?*
what do.PRES.1SG Monday.DEF?
'What do you do on Mondays?'
B: *Merg la film lunea / Lunea merg la film.*
go.PRES.1SG at movie Monday.DEF / Monday.DEF go.PRES.1SG at movie
'I go to the movies on Mondays.'

(13) *Merg la magazine azi / *aziul.*
go.PRES.1SG at shop today / *today.DEF
'I'm going shopping today.'

Similarly, yet another category of proper temporal names in Romanian, i.e., names of months, cannot take a definite determiner either. Since names of months cannot co-occur with the definite article in Romanian, the habitual-punctual distinction is excluded in this case, as illustrated by (14).

(14) *Merg la țară în mai / *maiul.*
go.1SG at countryside in May / *May.DEF
'I'm going to the countryside in May.'

As far as proper names of seasons are concerned, the first observation is that they are able to encode the punctual-habitual distinction by an alternation between a prepositional construction (15) and the emergence of the definite article (16). The prepositional construction in (15) encodes the punctual reading, while the definite article construction in (16) encodes the habitual reading.

(15) *Merg la țară la primăvară*
go.PRES.1SG at countryside at spring
'I'm going to the countryside this spring.'

(16) *Merg la țară primăvara.*
go.1SG at countryside spring.DEF
'I go to the countryside in spring.'

The cursory glance at the data indicates, therefore, that an analysis is needed to account for the association between the definite determiner accompanying names of days (and seasons) and the habitual interpretation.

3. Analysis

3.1. N-to-D raising with proper temporal names

This section aims at showing that, in Romanian proper temporal adverbials, like 'canonical' proper nouns in Italian (see Longobardi 1994, 2005, Alexiadou 2001, Tănase-Dogaru 2009, a.o.), are able to raise to D. Longobardi (1994) provided evidence for N-to-D raising in Romance languages, by mainly arguing that proper nouns raise to the D-position, the locus of referentiality. Starting from the examples in (17), Longobardi (1994) assumes that the proper noun needs to move from N^0 in order to fill in the empty D^0 position, thus crossing over the adjective/ possessive:

(17)
a. *Il mio Gianni ha finalmente telefonato* (Italian)
the my Gianni has finally called up

b. **Mio Gianni ha finalmente telefonato*
my Gianni has finally called up
c. *Gianni mio ha finalmente telefonato*
Gianni my has finally called up Longobardi (1994: 623)

In (17), the proper noun *Gianni* raises from N^0 in (17a) to D^0 in (17c), crossing over the possessive *mio* 'my' and incorporating the definite article. As shown by (17b), in the absence of the definite article, the proper noun cannot remain in N^0. In other words, the proper name either remains in N^0, while definiteness is expressed by the definite article in D^0, or it raises to D^0, where it incorporates definiteness.

In English, on the other hand, as illustrated by the contrast in (18a,b), proper nouns cannot raise to D^0. Therefore, by the logic of the 'only DPs as arguments' hypothesis, i.e., a nominal expression is an argument only if it is introduced by a category D, Longobardi (1994) assumes the existence of an empty determiner in the English example (18a), which is responsible for definiteness:

(18)
a. *Old John came in* (English)
b. **John old came in*. Longobardi (1994: 624)

The examples in (17) and (18) show that proper names move to D overtly in Italian, while English shows covert movement of the proper noun to D. In conclusion, Longobardi (1994) proposes a parametric variation for Italian and English: Italian substitutes N for D in overt syntax while English does the same at LF.

In Romanian proper nouns also bear definite articles (19a). As the examples in (19a-c) show, in contrast to languages like German (19b) and Greek (19c), which appear in the company of a proclitic definite article, Romanian proper nouns bear enclitic definite articles.

(19)
a. *Clujul e frumos. / I-am dat lui Ion cartea.*
Cluj.DEF is beautiful / CL.ACC.DAT.SG-have given DEF.DAT Ion book.DEF
'The city of Cluj is beautiful' / 'I gave Ion the book'
b. *Der Hans ist weggegangen.* (German)
DEF Hans is arrived
'Hans left.'
c. *O Kostis efuge.* (Greek)
DEF Kostis ran
'Kostis left.' Borer (2005: 85)

A very interesting phenomenon related to the fact that proper nouns have a complex syntactic structure is the behavior of animal names in Romanian. As shown in Tănase-Dogaru (2009: 139), animal names can get a proper noun interpretation when they appear without determiners, i.e., in D. In (20a), the noun is licensed as a common noun while in (20b) it is licensed as a proper noun.

(20)

a. *Viermele este scârbos. / Hiena este moartă*
worm.DEF is disgusting. / hyena.DEF is dead.
'The worm is disgusting.' / 'The hyena is dead'

b. *Vierme s-a supărat pe mine / Hienă a plecat în turneu.*
Worm is upset with me / hyena left in tour.
'Worm is upset with me'/ 'Hyena went on tour.' Tănase-Dogaru (2009: 139)

It is a familiar observation that, in Romanian, Differential Object Marking by means of *pe* selects type <e> entities, a reading which is reinforced by clitic-doubling. The contrasts in (21) clearly show that common nouns are licensed as proper nouns in the appropriate syntactic configurations:

(21)

a. *L-am văzut pe Vierme*
CL.ACC.MASC.SG-have.PRES.1SG seen DOM Worm
'I saw Worm.'

b. *Am văzut viermele.*
have.PRES.1SG seen worm.DEF
'I saw the worm.'

Having shown that proper nouns have a complex structure, and going back to the proper temporal names or adverbs in (7)-(8), repeated here for convenience as (22)-(23), I claim that this subset of proper nouns also has a complex syntactic structure.

(22) *Merg la magazine lunea.*
go.PRES.1SG at shop Monday.DEF
'I go shopping on Mondays.'

(23) *Merg la magazine luni.*
go.PRES.1SG at shop Monday
'I'm going shopping on Monday.'

As stated before, when the proper temporal name is accompanied by the definite determiner, the interpretation is necessarily habitual (22); when the proper temporal name is bare, the interpretation is necessarily punctual (23). It is the raising of the proper temporal name to D that gives rise to the punctual interpretation. The proper time adverb can move to the D position and it is spelt out in D, as illustrated in (24) (following Franco and Lorusso 2022). N-to-D movement triggers an individual-like reference, in a manner identical to the proper name movement to D in Longobardi (2005, 2008), the result of which is the interpretation of the event as punctual.

(24) [IP [VP merg [1SG [PP la magazin] [SpecDP luni … [NP luni]]]]
go.PRES.1SG at shop Monday (see Franco and Lorusso 2020: 5)

This type of analysis is in line with other observations in the literature on proper temporal names (see, for instance Anderson 2003, 2004, 2007 and Langendonck (2007). Anderson (2003: 367) claims that proper temporal names, or *calendrical* names are hybrid, in the sense that they

are deictically restricted names that can also be used as count nouns (see Anderson 2003: 367). Langendonck (2007: 223) notices the same ambiguity in English examples such as (25a,b):

(25)
a. *June is (usually) a hot month.*
b. *June was hot.*

(25a) shows a 'recursive (cyclic, generic, habitual)' interpretation of the month name *June*, which indicates the spelling out of *June* in N. (25b) exemplifies a non-recursive, more 'prototypical' deictic reading of the proper name, which shows that the proper name has moved and is spelled out in D (as in (24)). Since English has bare proper names, the punctual vs. habitual distinction can only be encoded by means different aspectual make-ups, i.e. the present in (25a) vs. the past in (25b), unlike languages like Italian or Romanian, which have proper names surfacing in the company of definite articles, proper temporal names included. As originally observed by Franco and Lorusso (2022: 7), the situation in (25a,b) is related to the punctual vs. habitual (or bounded vs. unbounded) value of Italian proper temporal names. As argued throughout in this paper, the same punctual vs. habitual value can be ascribed to Romanian proper temporal names. As Franco and Lorusso put it, "there must be a way to turn a proper name into a countable entity in order to obtain a habitual/iterative value for the event described by means of proper temporal adverbials" (Franco and Lorusso 2022: 7).

This section has shown that proper temporal names can move to D, in a manner similar to the way proper names undergo N-to-D movement. The reflex of this syntactic movement is the punctual interpretation. The next section shows that the syntactic structure of (temporal) proper names contains a classifier; this classifier is overt in the adjective – proper name order, i.e., in the absence of N-to-D movement.

3.2. On classifiers and proper (temporal) names

The main focus of this section is to discuss the complex syntactic structure of proper (temporal) names, illustrated in (26)-(29). (26) illustrates the punctual reading of the proper temporal name *marți* 'Tuesday', while (27) illustrates N-to-D movement across the adjective *viitoarea* 'next' resulting in the punctual interpretation, mainly because of the meaning of the adjective 'next'. (28) illustrates an interesting property of the proper temporal name accompanied by an adjective like *viitoare* 'next', which triggers the punctual reading, namely the presence of the adjective triggers the overt realization of a classifier *zi* 'day', which would otherwise remain silent. (29) shows that the presence of an evaluative adjective like *minunata* 'wonderful' also mandatorily triggers the overt realization of the classifier.

(26) *Merg* *la film* *marți.*
go.PRES.1SG to movie Tuesday
'I'm going to the movies on Tuesday.'

(27) *Merg* *la film* *marțea* *viitoare.*
go.PRES.1SG to movie Tuesday.DEF next
'I'm going to the movies next Tuesday.'

(28) *Merg la film viitoarea *(**zi** de) marți.*
go.PRES.1SG to movie next.DEF (day of) Tuesday
'I'm going to the movies next Tuesday.'

(29) *minunata *(**zi** de) marți*
wonderful.DEF (day of) Tuesday
'the wonderful Tuesday'

The main claim of the section is that this syntactic structure of proper (temporal) names contains a classifier *zi* 'day', which is silent whenever there is N-to-D movement, as shown by (26) and (27); the classifier is, however, obligatorily overt when the proper temporal name stays in N, as illustrated by (28) and (29).

3.2.1. What's in a name? A classifier

Starting from an investigation of nominal constructions of the type in (30)[4], which go by many names in the literature, such as restrictive appositives, qualificational nouns, descriptive proper names, complex proper names, qualitative classifiers (van Riemsdijk 1998, Löbel 2001, Cornilescu 2007, Tănase-Dogaru 2011, 2013), this section shows that this type of nominal constructions contains an overt qualitative classifier in the syntactic make-up (Cornilescu 2007, Tănase-Dogaru 2013).

(30)

a. *Profesorul Ionescu*
professor.DEF Ionescu
'Professor Ionescu'

b. *Regina Elizabeta*
queen.DEF Elisabeth
'Queen Elizabeth'

c. *Mătuşa Tamara*
aunt.DEF Tamara
'aunt Tamara'

d. *Prinţul Carol*
prince.DEF Charles
'Prince Charles'

e. *Oraşul Iaşi*
city.DEF Iasi
'the city of Iaşi'

f. *Strada Paris*
street.DEF Paris
'Paris street'

Cornilescu (2007: 63)

Complex or descriptive proper nouns are proper nouns formed of a common noun + proper noun. The complex or descriptive noun designates a social role (kinship, profession, institutional role), or a sort of place (city, street, river, village, etc.), some other entity (a theatre, a planet, etc.) (see Cornilescu 2007).

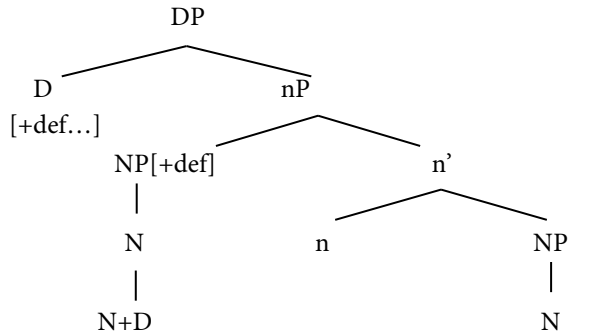
[4] I will refer to this type of construction with the term 'complex proper noun' or 'descriptive proper noun'.

Therefore, in (30), the functional structure of the proper nouns includes not only a D [+def, +phi, + Person] but also a noun designating a social role, a profession, a name of place, etc., which functions as a (qualitative) classifier. This conception reflects the intuition that a complete understanding of a proper noun requires identifying the kind of entity that it names (Cornilescu 2007: 61).

The role of the common noun is that of classifying entities, indicating the kind of entity the proper name denotes, as in *Professor Smith*. While for simple proper nouns, the proper noun itself checks the classifier feature, by Move or Agree, for complex or descriptive proper nouns, the descriptive common noun merges as the specifier of the nominal-class head, since the feature of this n-head is one of the features of the common noun. With complex or descriptive proper nouns, the classifier head is overt (Cornilescu 2007: 63).

The structure of a Romanian complex or descriptive proper noun is given in (31) (from Cornilescu 2007: 75), where the proper noun is too low to check [+def, +phi + person], so the descriptive NP must be definite, and checks the D[+def] feature.

(31)



Cornilescu (2007: 75)

3.2.2. Classifiers and names of months and years

This section shows that 'names' of years, months and hours also behave like complex/ descriptive proper nouns containing (silent or overt) classifiers (see also Tănase-Dogaru 2009, 2013). In Romanian, there is an interesting variation between constructions like (32a) and (32b):

(32)

a. *Benedict s-a născut în anul 480.*
Benedict REFL-has born in year.DEF 480
'Benedict was born in the year 480.' (CoRoLa[5])

b. *O a doua invazie a urmat în 480.*
a second invasion has followed in 480.
'A second invasion followed in 480.' (CoRoLa[6])

Although years prior to 1000 are better with the overt classifier 'anul' / 'year.DEF', the same variation can be seen with constructions involving more recent years (33a-b):

[5] https://korap.racai.ro/?q=%C3%AEn+anul+480&ql=cosmas2&cutoff=1, accessed September 27, 2021

[6] https://korap.racai.ro/?q=%C3%AEn+480&ql=cosmas2&cutoff=1, accessed September 27, 2021

(33)

a. *Slavici îl urmează în anul 1880.*
Slavici CL.ACC.MASC.SG follows in year.DEF 1880.
'Slavici follows him in the year 1880.'

b. *Oficial, în 1880, în regiune trăiau 241 de persoane.* (CoRoLa[7])
officially, in 1880, in region live.IMPERF.3.PL 241 of persons
'Officially, in 1880, 241 persons inhabited the region.' (CoRoLa[8])

The examples clearly show that, when referring to years, there is variation between the silent and the overt classifier 'anul' / 'year.DEF', which makes years very similar to descriptive proper names.

However, when there is an adjective involved in the syntactic structure of complex proper names of years, the classifier is overt (34). The same phenomenon was illustrated for proper temporal names of days, in the sense that the adjective forces the spelling out of the classifier (see (27) and (28) above):

(34) *groaznicul an 1946*[9]
horrible.DEF year 1946
'the horrible year of 1946'

The same variation can be noticed with names of months. The examples in (34a) and (34b) show that, in the case of names of months, the classifier 'luna' / 'month.DEF' is either silent (34a) or overt (34b).

(34)

a. *în aprilie devine din nou un om liber.*
in April become.3.SG again a man free
'in April he becomes again a free man' (CoRoLa[10])

b. *în luna aprilie au loc expoziții și festivaluri*
in month.DEF April have.3.PL place exhibitions and festivals
'In the month of April exhibitions and festivals take place.' (CoRoLa[11])

The presence of an adjective which agrees in feminine gender with the overt classifier *luna* 'month', which has feminine gender features, triggers the spelling out of the classifier (35a). In contrast, (35b) and (35c) show that, when the classifier is silent, it has (default) masculine gender, so that the adjective agrees in masculine gender with the silent classifier.

---

[7] https://korap.racai.ro/?q=%C3%AEn+anul+1880&ql=cosmas2&cutoff=1, accessed September 27, 2021

[8] https://korap.racai.ro/?q=%C3%AEn+anul+1880&ql=cosmas2&cutoff=1, accessed September 27, 2021

[9] An anonymous reviewer points out that the structure *groaznicul 1946* 'horrible.DEF 1946' is also possible in Romanian, which I definitely agree with. However, my (naïve) informants prefer the variant with the overt classifier *anul* 'year.DEF'.

In close connection with this matter, the same preference is observed in *groaznicii ani '80* 'terrible.PL.DEF years '80' over *??groaznicii '80* 'terrible.PL.DEF '80s'. This reflects the intuition in Cornilescu (2007: 61) that a "complete understanding of a proper name requires identifying the kind of entity that it names".

[10] https://korap.racai.ro/?q=%C3%AEn+aprilie&ql=cosmas2&cutoff=1, accessed September 27, 2021

[11] https://korap.racai.ro/?q=%C3%AEn+luna+aprilie&ql=cosmas2&cutoff=1, accessed September 27, 2021

(35)

a. *minunata* *lună* *aprilie*
wonderful.FEM.DEF month.FEM April
'the wonderful month of April'

b. *minunatul* *aprilie* = *minunatul* *MONTH* *aprilie*
wonderful.MASC.DEF April = wonderful. MASC.DEF MONTH.MASC.DEF April
'the wonderful April'

c. **minunata* *aprilie*
wonderful.FEM.DEF April
'the wonderful April'

This section has argued that 'names' of years and months behave syntactically in a fashion parallel to descriptive proper names of the type *the planet Venus* (van Riemsdijk 1998). These constructions possess a classifier in their make-up, a classifier that is either silent or overt. The spelling out of the classifier is related to the presence of an adjective in the structure. The next section takes a look at expressions involving the classifier *hour*, which in turn, is argued to be part of the syntactic make-up of expressions telling the time.

3.2.3. Classifiers and names of hours

Kayne (2005: 258) notes that in the English example (35), HOUR acts as a silent classifier:

(35) *It's six.*
It's six HOUR.

As for Romance languages, Kayne (2005) notes that French (36) differs from Italian (37) with respect to time, in that the classifier *heures* / 'hours' must be overt in French:

(36) *Il est six heures.*
It is six hours
'It's six o'clock.'

(37) *Sono le sei.*
are the six
'It's six o'clock.'

In Italian, the corresponding noun can be present, although that is less usual:

(38) *Sono le ore sei.*
are the hours six
'It's six o'clock.'

Kayne (2005: 259) argues that the obligatory presence of the classifier in French is related to the presence of the definite article *le* in Italian (38) versus its absence in French (36). I take this to mean that the presence of overt lexical material in the D layer is able to license silent classifiers, i.e. the case of Italian, while the absence of the D layer correlates with the overt classifiers, i.e. the case of French.

Romanian exhibits the contrast in (39a,b), which shows that the classifier may be either silent or overt. However, the most common way of telling the time is (39a), where the classifier is silent.[12]

(39)
a. *E șase.*
is six
'It's six o'clock.'
b. *E ora șase.*
is hour six
'It's six o'clock.'

The classifier HOUR is overt in examples such as (40):

(40)
a. *De la ora 6 dimineața am plecat.*
of at hour.DEF 6 morning.DEF have.1.SG left.
'I left at 6 o'clock in the morning.' (CoRoLa[13])
b. *Era dimineața la ora 6.*
Was morning.DEF at hour.DEF 6
'It was six o'clock in the morning.' (CoRoLa[14])

Like Italian, Romanian also has the variant in (41), which is, however, perceived as bookish and obsolete. In this case, the overt classifier surfaces in the plural.

(41) *pe la orele 6 am făcut un mic popas*
on at hours.DEF 6 have.1.SG made a small stop
'at about 6 o'clock we had a short break' (CoRoLa[15])

Therefore, it seems safe to assume that, in telling the time, classifiers are also present, whether silent or overt.

The presence of an adjective in the complex structure of these proper names triggers the overt realization of the classifier (42):

(42)
a. *groaznica* *oră* *6*
horrible.FEM.DEF hour 6
'the horrible 6'
b. **groaznica / groaznicul* *6*
horrible.FEM.DEF / horrible.MASC.DEF 6
'the horrible 6'

Section 3.2 has shown that the structure of complex proper names, including names of social roles, names of places, and names of years, months and hours, contains a classifier. In the case

[12] The preference of (39a) over (39b) may be related to the clash between a plural feature of the cardinal for 6 and the singular feature of the classifier *ora* 'hour'. I leave this matter to further research.

[13] https://korap.racai.ro/?q=ora+6&ql=cosmas2&cutoff=1, accessed September 27, 2021

[14] https://korap.racai.ro/?q=ora+6&ql=cosmas2&cutoff=1, accessed September 27, 2021

[15] https://korap.racai.ro/?q=orele+6&ql=cosmas2&cutoff=1, accessed September 27, 2021

of proper names of years, months, and hours, the classifier may be silent or overt. The classifier, however, is (necessarily) overt when the noun is accompanied by an adjective.

4. The syntactic structure and the role of *de/di* 'of'

This section capitalizes on what has been said so far with respect to the syntactic structure of complex proper names, including proper temporal names, and proposes a syntactic structure of the type classifier-noun for proper temporal adverbs. It was shown that the structure of proper (temporal) names contains a classifier that can be silent or overt. It seems natural to assume that the underlying representation of (43a) should be (43b):

(43)
a. *marți*
Tuesday
b. *ZIUA DE marți*
DAY of Tuesday

The same can be assumed about Italian, as shown by the contrast between (44a) and (44b), the difference being that *di* 'of' is overt:

(44)
a. *martedì*
Tuesday
b. *GIORNO DI martedì*
DAY OF Tuesday

Since one of the pre-conditions for the postulation of silent elements is their attestation in the overt form, (45a,b) show that the classifier DAY can also be overt:

(45)
a. *De ce este ziua de marți atât de încărcată de superstiții*[16] (Romanian)
of what is day of Tuesday so loaded of superstitions
'why is the day of Tuesday so loaded with superstitions'
b. *Ma un giorno di lunédi capirai* […][17] (Italian)
but a day of Monday understand.FUT.2SG
'One Monday you will understand'

In order to articulate a syntactic structure for proper temporal names, the role of *de/di* 'of' needs to be determined. In Italian (see Franco and Lorusso 2022), the prepositional element *di* 'of' appears with proper temporal names of days and lends the example a habitual reading (46), a reading also available with the definite article (47):

[16] https://www.libertateapentrufemei.ro/superstitii/de-ce-este-ziua-de-marti-atat-de-incarcata-de-superstitii-212297

[17] https://www.spreaker.com/user/webradio11-11/il-ritorno-di-adriano-celentano-ma-un-gi

(46) *vado al cinema di lunedì* = habitual reading
go.pres.1SG to.the cinema of Monday
'I (usually) go to the cinema on Mondays.'

(47) *vado al cinema il lunedì* = habitual reading
go.pres.1SG to.the cinema the Monday
'I (usually) go to the cinema on Mondays.'

Manzini and Savoia (2011) and Franco and Lorusso (2020) analyze the *di* element as a genitive-assigning preposition, different from the partitive marker *dei,* but still involving an 'inclusion/sub-set' relation.

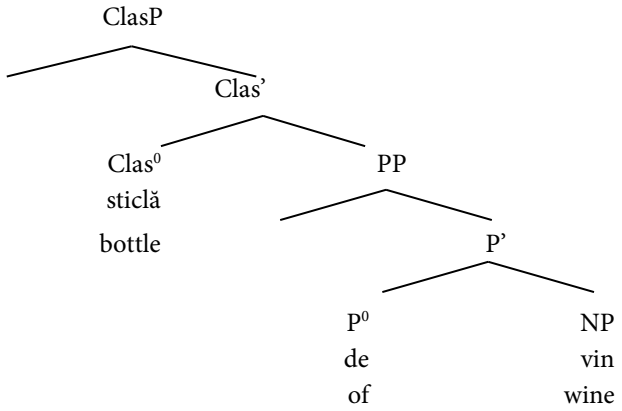
I claim that the same happens in Romanian. Whenever the classifier is overt, as in (48), the *de* element is a partitive-genitive preposition. Moreover, the syntactic structure of such constructions parallels the one present with pseudo-partitive constructions generally (49) (Tănase-Dogaru 2008, 2009, 2017).

(48) *ziua de marți*
day of Tuesday

(49) *sticlă de vin*
bottle of wine

Pseudo-partitive constructions are seen as single multi-headed extended projections, as shown in (50). N_1 in pseudo-partitives is a semi-lexical or functional element which behaves as a classifier, turning pseudo-partitives into Classifier Phrases (Tănase-Dogaru 2008, 2009). (50) gives the syntactic representation of pseudo-partitives, where *de* 'of' is a partitive marker, assigning N_2, i.e., the lexical head, abstract genitive case (Tănase-Dogaru 2008, following van Riemsdijk 1998, Vos 1999).

(50)



Tănase-Dogaru (2008: 315)

Classifiers project universally in the functional architecture of nominal phrases. The head of the ClasP may be filled with lexical material – as in the case of pseudo-partitive constructions – or with silent material (see Kayne 2005, van Riemsdijk 2005, Tănase-Dogaru 2008, 2009). For instance, in the pair (51a,b), (51b) contains the silent noun NUMBER acting as a classifier:

(51)

a. *Ce case au unii!*
what houses have some (people)
'some have such big/beautiful houses!'

b. *Ce de case au unii!*
what of houses have some (people)
'some have so many houses!' Tănase-Dogaru (2008: 317)

(51a) is an exclamation about some salient property of houses, for example their being large or beautiful; on the other hand, (51b) can only be an exclamation about the relatively large number of the houses in question.

The syntactic structure of proper temporal names also contains a classifier, which may be silent, i.e., DAY or overt, i.e., *zi* in Romanian, *giorno* in Italian. In what follows, I will exemplify the syntactic structures of the Italian and Romanian proper temporal name constructions. I will start with the Italian data, simplified for the current purposes as (52a-c).

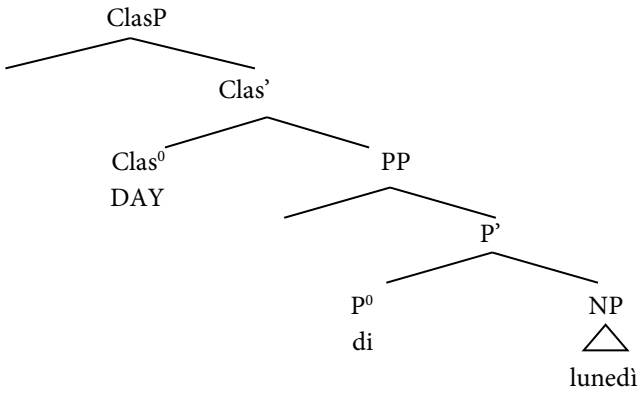
(52)

a. *di lunedì* (habitual reading) = DAY di lunedì
b. *il lunedì* (habitual reading) = il DAY OF lunedì
c. *lunedì* (punctual interpretation) = DAY OF lunedì

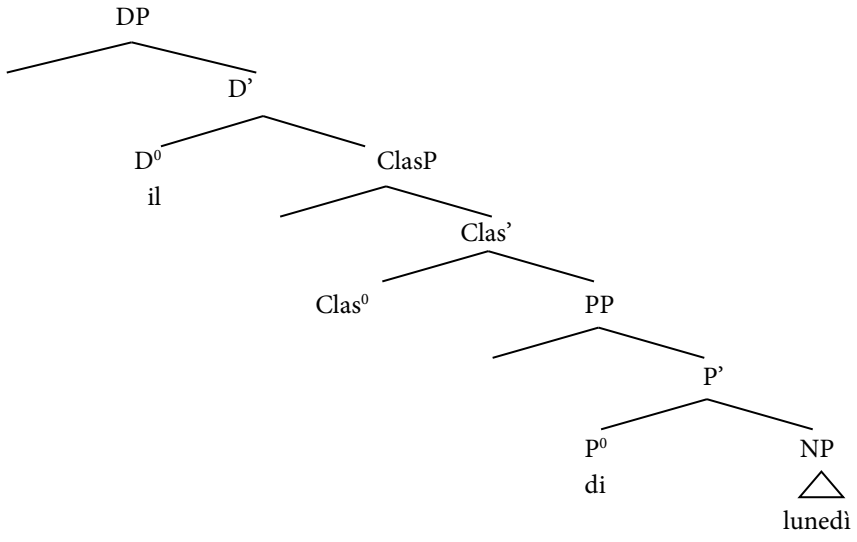
The corresponding syntactic structures are given in (53a-c).

(53)

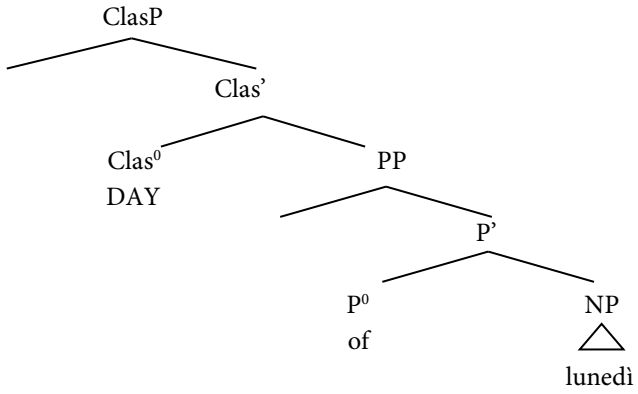
a.



b.



c.



The Romanian data are shown again in (54a,b), with their corresponding syntactic structures in (55a,b):

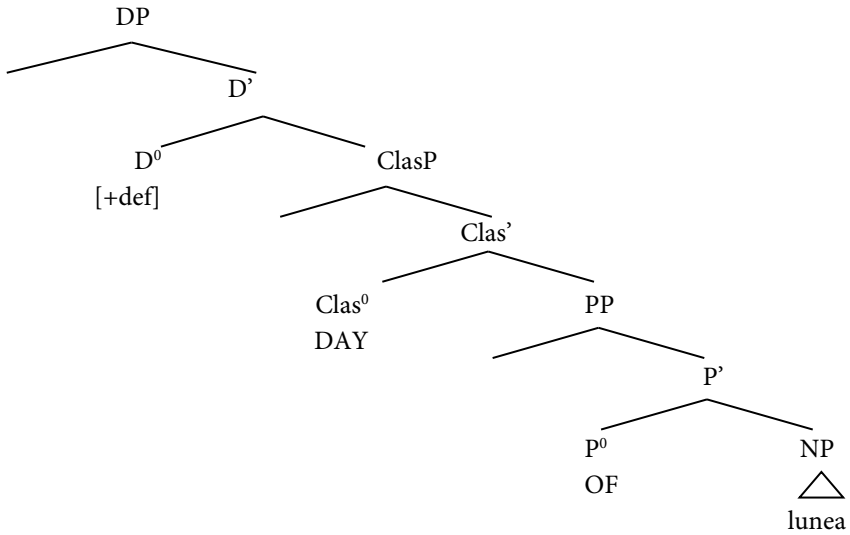
(54)

a. *lunea* (habitual reading) = DAY OF lunea

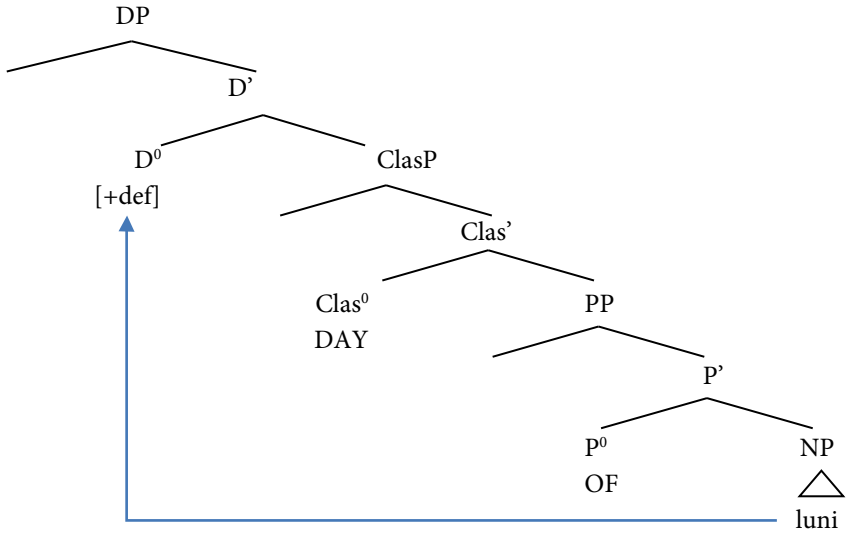
b. *luni* (punctual reading) = DAY OF luni

(55)

a.



b.



In (55a), lunea 'Monday.DEF[18]' is generated in N, a position it does not leave, it is the lack of movement which accounts for the habitual reading. (55b) illustrates N-raising of the proper temporal name to D.

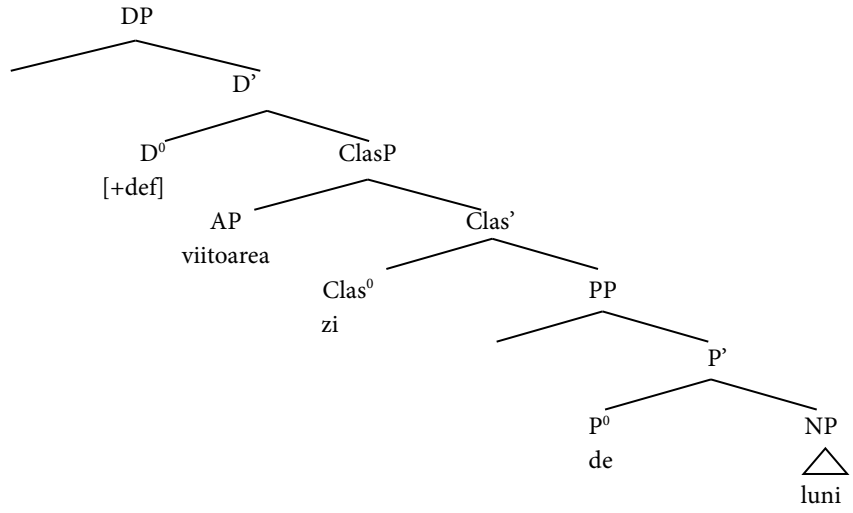
This type of movement is visible in examples (56a,b), with the corresponding syntactic structures in (57a,b). In (56a), the emergence of the adjective triggers the overt realization of the classifier DAY, while in (56b), there is N-movement to D, and the classifier remains silent. The adjective in (56a) modifies the classifier, a relation which triggers the overt realization of the classifier.

(56)

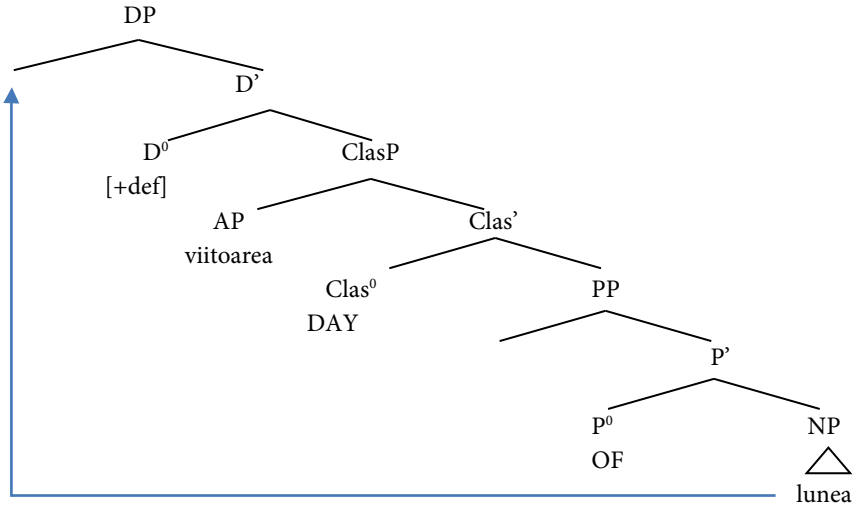
a.	*viitoarea*	*zi*	*de luni*
	next.DEF	day	of Monday
b.	*lunea*	*viitoare*	
	Monday.DEF	next	

(57)

a.



b.



[18] I follow Nicolae (2013) and Cornilescu (2016) in arguing that Romanian has a suffixal definite article. In a simple structure where the DP consists only of a definite noun, the uninterpretable definite feature of the noun values the interpretable definite feature of the D Probe by means of the operation Agree (Cornilescu 2016: 14). This means that a DP like *lunea* 'Monday.the' in (55a) remains in N and undergoes Agree with the definite feature in D.

In (57a), *luni* 'Monday' is generated in the N layer, in a typical classifier-noun sequence of the type discussed in section 4. In (57b), *lunea* 'Monday.the' undergoes movement to SpecDP, crossing over the adjective *viitoare* 'next', in a manner similar to proper names in Italian (see section 3.1)

This section has shown that the syntactic structure of proper temporal names is complex, in the sense that it contains a classifier, which may be silent or overt. The overt realization of the classifier is triggered by the presence of adjectives, while the prepositional element *de/di* 'of' is a partitive-genitive marker.

5. Conclusions and further research

The paper has started from the empirical observation that proper temporal names in Romanian may have punctual or a habitual interpretation. These interpretations are associated with the absence or presence, respectively, of the definite article. With bare proper temporal names, N-to-D movement triggers individual-like reference, which, in turn, explains why the event is interpreted as punctual. An important finding of the paper is related to the complex syntactic structure of proper temporal names. This syntactic structure contains classifiers, which may be silent or overt. The silence of the classifiers is related to the phenomenon of N-raising. I leave for further research a finer-grained parameterization of Italian and Romanian with respect to the silence of the prepositional element *di/de* 'of'.

References

Alexiadou, A. 2001. *Functional structure in nominals*. Amsterdam/Philadelphia: John Benjamins.

Anderson, J. M. 2003. On the structure of names. *Folia Linguistica* 37: 347–398.

Anderson, J. M. 2004. On the grammatical status of names. *Language* 80: 435–474.

Anderson, J. M. 2007. *The grammar of names*. (Oxford Linguistics). Oxford: Oxford University Press.

Borer, H. 2005. *In Name Only*. Oxford: Oxford University Press.

Cornilescu, A. 2007. On classifiers and proper names. *Bucharest Working Papers in Linguistics* 9 (1): 62–75.

Cornilescu, A. 2016. The low definite article and the evolution of the Romanian DP. *Generative Grammar in Geneva* 9: 3–26.

Franco, L., and P. Lorusso. 2020. On the interpretation of proper temporal adverbs in Italian: The role of P and D. Talk delivered at Talk delivered at Going Romance, November 2020. Handout available at https://www.sfl.cnrs.fr/going-romance-34-special-session-programme.

Franco, L., and P. Lorusso. 2022. On the interpretation of proper temporal adverbs in Italian: The role of P and D. Isogloss. *A journal on variation of Romance and Iberian languages* 8(2): 1–18.

Kayne, R. S. 2005. *Movement and Silence*. Oxford: Oxford University Press.

Langendonck, van W. 2007. *Theory and typology of proper names*. Berlin: Mouton De Gruyter.

Longobardi, G. 1994. Reference and Proper Names: a Theory of N- Movement in Syntax and Logical Form. *Linguistic Inquiry* 25(4): 609–665.

Longobardi, G. 2005. Toward a Unified Grammar of Reference. *Zeitschrift für Sprachwissenschaft* 24: 5–44.

Löbel, E. 2001. Classifiers and semi-lexicality: Functional and semantic selection. In N. Corver, and H. van Riemsdijk (eds.), *Semi-lexical Categories. The Function of Content Words and the Content of Function Words*, 223–271. Berlin: Mouton de Gruyter.

Manzini, M. R., and L. M. Savoia. 2011. Reducing 'case' to denotational primitives. *Linguistic Variation* 11: 76–120.

Nicolae, A. 2013. Notă de sintaxă comparată: parametrul [+definit] în sintaxa grupului nominal românesc. *Limba Română* 62(2): 186–209.

Riemsdijk, H. C. van. 1998. Categorial feature magnetism: The endocentricity and distribution of projections. *Journal of Comparative Germanic Linguistics* 2: 1–48

Riemsdijk, H. C. van. 2005. Silent nouns and the spurious indefinite article in Dutch. In M. Vulchanova, and T. A. Åfarli (eds.), *Grammar and Beyond: Essays in honour of Lars Hellan*, 163–178. Oslo: Novus Press.

Tănase-Dogaru, Mi. 2008. Pseudo-partitives and (Silent) Classifiers in Romanian. In S. Blaho, C. Constantinescu, and E. Schoorlemmer (ed.), *Proceedings of ConSole XV*, 295–320. Leiden: Sole.

Tănase-Dogaru, M. 2009. The Category of Number. *București: Editura Universității din București* XIII (1): 151–167.

Tănase-Dogaru, M. 2011. Diagnostic Criteria for Semi-Lexicality. Bucharest Working Papers in Linguistics.

Tănase-Dogaru, M. 2013. Classifiers of quantity and quality in Romanian. *Bucharest Working Papers in Linguistics*, vol. XV, nr. 2: 77–90.

Tănase-Dogaru, M. 2017. *Partitive constructions. The Wiley Blackwell Companion to Syntax*. Second Edition, edited by Martin Everaert and Henk van Riemsdijk. https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/9781118358733.wbsyncom017

Vos, Riet. 1999. A Grammar of Partitive Constructions. PhD Diss: Tilburg University.

LingBaW *Linguistics Beyond And Within 8 (2022), 200-217*

The morphology of complex numerals: A cross-linguistic study

Lukáš Žoha, Marcin Wągiel and Pavel Caha
Masaryk University, Brno, Czech Republic

Abstract
Complex numerals are numerals composed of two or more numeral roots, e.g., *three hundred five*. Complex numerals fall into two classes called additive (e.g., *twenty-three* = 20 + 3) and multiplicative (e.g., *three hundred* = 3 × 100). There are two possible approaches to capturing their structure. Analysis A (e.g., He 2015) says that complex numerals form a constituent that quantifies over entities denoted by the noun. Analysis B (e.g., Ionin and Matushansky 2018) says that each numeral independently combines with the expression denoting counted entities. This article investigates the morphology of complex numerals in a sample of 17 diverse languages to determine which of these analyses (if any) is more accurate. Our goal is to lay out the patterns and discuss how well they fit with these theories. Our preliminary conclusion is that both structures should be allowed based on the data in our sample, though structures adhering to Analysis A (the complex numeral is a constituent) seem to be more common than the other type.

Keywords: complex numerals; counting; morphology; syntax; linguistic typology

1. Introduction

From the perspective of their composition, numerals can be divided into simple numerals and complex numerals. In (1a), there is an example of a simple numeral containing just a single numeral root. In (1b) and (1c), there are examples of complex numerals of which the simple numeral *three* is a part.[1]

(1) English
a. ***three***
b. *one hundred* ***three***
c. ***three*** *hundred*

[1] The term complex numerals currently has different uses in the literature. In Wągiel and Caha (2021), the term refers to numerals that are morphologically complex, e.g., the German *ein-s* 'one' has two morphemes. Here we use it differently, namely as referring to numerals that contain more simple numerals, as in (1b-c). This latter use of the term 'complex' is also found in Ionin and Matushansky (2018).

Complex numerals can be further divided into two groups, based on how we arrive at their meaning. If addition is involved, as in (1b), the complex numeral is called additive (100 + 3). If multiplication is involved (3 × 100), the complex numeral is called multiplicative, see (1c). (Addition and multiplication can of course combine within a single numeral with more than two elements.)

A boarder-line case is represented by numerals such as *six-ty*, where clearly, we have the numeral *six* and an additional morpheme *-ty*. Even though *-ty* is not a free-standing numeral, it is a morpheme with a constant contribution across a range of cases (*seven-ty*, *nine-ty*, etc.). We consider these cases as relevant for our study because we mainly focus on the form of the multiplier (*six, seven*) rather than on the form of the irregular base (*-ty*).

There are two main approaches to the structure of complex numerals. Under Analysis A (e.g., He 2015, Cinque 2021, He and Her to appear), complex numerals form a constituent that combines with the noun and quantifies over the entities denoted by the noun. This is shown in (2a,b) for additive and multiplicative complex numerals, respectively. In both cases, the complex numeral is formed first and then it is added to the noun as a constituent.

(2) Analysis A: Complex numerals as constituents
a. [twenty + three] books
b. [three × hundred] books

Under Analysis B (e.g., Ionin and Matushansky 2006, 2018), each numeral independently combines with the expression denoting counted entities. For additive numerals, Analysis B leads to a structure like (3a). At the beginning, there are two phrases with two nouns: *twenty books* and *three books*. The complex numeral *twenty-three* is created from these two phrases by eliminating the noun from the phrase *twenty books* (either by right node raising or phonological deletion).

(3) Analysis B: Complex numerals as non-constituents
a. [twenty books] + [three books]
b. [three × [hundred books]]

For multiplicative numerals, Analysis B leads to a structure like (3b), where the numeral *three* counts the number of [hundred books]. In neither of these structures do the two numeral roots form a constituent in the underlying structure to the exclusion of the modified noun.

There are several empirical tests that allow one to distinguish between the two analyses (see He 2015, Ionin and Matushansky 2018, He and Her to appear for a recent discussion). In this paper, we try to differentiate the two hypothetical structures by a novel test that has not been used in the literature so far. The idea is that in at least some languages, the two structures can be distinguished on morphological grounds by looking at the morphology of simple numerals inside complex numerals.

This new diagnostic builds on the observation by Wągiel and Caha (2020, 2021) that languages may morphologically distinguish between two kinds of numerals, which we call here abstract- and object-counting numerals (following the terminology in Wągiel and Caha).

Abstract-counting numerals refer to an abstract number concept (Bultinck 2005, Rothstein 2017). They are therefore used in statements expressing mathematical properties, as in (4a), or arithmetical equations, as in (4b); the relevant numerals are highlighted. Often (though not always), the same numerals are also used for counting in a sequence, see (4c). (The counting sequence is grouped with the arithmetical uses, e.g., in Tatsumi 2021.) The second type of numerals that we need to recognize are object-counting numerals. These are forms used inside NPs for counting objects, as in (4d).

(4) English
a. ***Three*** *is a prime number.*
b. *Two plus* ***three*** *equals five.*
c. *one, two,* ***three,*** *…*
d. *There are* ***three*** *cats in the room.*

In English there is no morphological distinction between abstract-counting and object-counting numerals. In some languages, however, these numerals have different forms. For instance, the Mandarin numeral 'two' is *èr* for abstract counting, and *liǎng* + classifier for object counting (He 2015). We shall return to this example in Section §3.3.

Another language that distinguishes abstract- from object-counting numerals is Javanese. As we show in (5), in this language the numeral 'five' has three different forms, ordered in terms of morphological complexity.

(5) Javanese 'five' (Robson 1992: 75–76)
a. ***ma***
b. *li-**ma***
c. *li-**ma**-ng*

The numeral *ma* in (5a) can only be used when numerals are recited in the counting sequence, and we, therefore, consider it an abstract-counting numeral. The numeral *lima* in (5b) is ambiguous like English 'five,' and it can be used both in the counting sequence and as an object-counting numeral. This latter use is illustrated in (6a).

(6) Javanese (Robson 1992: 75–77)
a. *jeruk* ***lima***
orange five
'five oranges'
b. ***limang*** *rupiah*
five rupiah
'five rupiahs'

The numeral *limang* can only be used for object counting. In (6b), it is used for counting currency, and thus we classify it as an object-counting numeral.

The idea that we follow in this paper is that in languages where we find a difference between object- and abstract-counting numerals, it is relevant to look at which of these numerals is used in object-counting complex numerals. To see why, consider again the two analyses, repeated for convenience in (7)–(8):

(7) Analysis A: Complex numerals as constituents
a. [twenty + three] books
b. [three × hundred] books

(8) Analysis B: Complex numerals as non-constituents
a. [twenty books] + [three books]
b. [three × [hundred books]]

Starting with Analysis B, we expect that each numeral inside the complex numeral is an object-counting numeral, because it is associated to a particular entity it quantifies over. If even one abstract-counting numeral is found inside a complex numeral, this would be unexpected.

On the other hand, Analysis A makes different predictions. For additive numerals, it predicts that at least one numeral inside the complex numeral is in the abstract-counting shape. The reason is the following. According to the simplest version of Analysis A, the numeral is created first through an abstract arithmetic operation (20+3), and this requires abstract-counting numerals. Only after the numeral is formed, it is used to quantify over objects. This hypothesis therefore predicts that inside additive complex numerals, abstract-counting numerals should be used.

However, it is not expected that both numerical elements have to be in the abstract-counting form. This is because morphemes indicating the object-counting function may attach to the complex numeral as a whole. In such cases, we expect that morphology conveying the object-counting function will affect the edges of the complex numeral (e.g., by a phrasal affix), but not its internal composition.

Turning now to multiplicative complex numerals, Analysis A also leads us to expect abstract-counting numerals inside the complex. However, it is not *a priori* clear whether the constituency in (7b) necessarily leads to the obligatory occurrence of abstract-counting numerals; it could be that the numeral *three* counts *hundreds* in the same way as it counts other objects. If so, the constituency could be [[*three hundred*] *books*] even if *three* were an object-counting numeral. However, clearly, if *three* were in the abstract-counting shape, this would support the constituency in (7b).

This paper presents relevant examples from a database of 17 languages that distinguish abstract-counting and object-counting numerals in at least a subpart of their inventory. Our goal is to lay out the patterns according to the expectations just described. We leave it open as to whether the respective analyses may be combined with some additional assumptions that would change the basic expectations emanating from the different structures.

2. Possible patterns of complex numerals

Based on the reasoning in the preceding section, we have constructed a sample of 17 languages that make a distinction between object- and abstract-counting numerals in at least a part of their inventory of numerals. The languages come from twelve different language families:

(9) Bangla, Irish, Romanian (Indo-European), Javanese, Puyuma (Austronesian), Swahili, Zulu (Niger-Congo), Mandarin, Shuhi (Sino-Tibetan), Telugu (Dravidian), Thai (Kra-Dai), Huehuetla Tepehua (Totonacan), Maltese (Afroasiatic), Palikur (Arawakan), Vietnamese (Austroasiatic), Japanese and Korean.

In this section, we discuss the patterns we found. Recall that (nothing else said), Analysis B predicts that complex numerals contain object-counting numerals only. We discuss such languages in Section §2.1 (there is only one such language in our sample, namely Irish). In Section §2.2, we discuss languages where complex numerals always contain at least one abstract-counting numeral. There are seven languages of this type in our sample: Mandarin, Korean, Japanese, Thai, Vietnamese, Telugu, and Bangla. Recall that these languages point in the direction of Analysis A. In Section §2.3, we discuss languages which mix these two patterns, i.e., where some complex numerals contain only object-counting numerals, while others contain at least one abstract-counting numeral (Maltese, Puyuma, Romanian, Javanese). Finally, in Section §2.4, we discuss languages with patterns that cannot be subsumed under the previous types for various reasons (Shuhi, Huehuetla Tepehua, Palikur).

Based on these results, we shall conclude that both types of structures seem to be attested in natural languages (a conclusion suggested also in Wągiel and Caha 2021: fn 12). We note, however, that our approach is mainly descriptive, taking the data at face value without any attempts to reinterpret them *per force* one way or another. At the same time, we do not want to preclude the option that this may be possible (see, e.g., Ionin and Matushansky 2018: 126–139).

2.1. Patterns of complex numerals with object-counting numerals only

This type of pattern can only be found in one language of the sample, namely Irish. The language belongs to Indo-European language family (Celtic), and it is spoken in Ireland. In Irish, simple abstract-counting numerals from 'one' to 'ten' are always preceded by an unstressed *a* (Dylon and Ó Cróinín 1961: 64). We give an example in (10a), glossing *a* as NBR (number). In (10b), we illustrate the fact that object-counting numerals lack the initial *a*.

(10) Irish ((a) is adapted from Stenson 2020: Ch. 20.1; Dylon and Ó Cróinín 1961: 63, 137)

a. ***A*** ***dó*** *agus* ***a*** ***dó*** *sin* ***a*** ***ceathair.***
NBR two plus NBR two are NBR four
'Two plus two are four.'

b. ***trí*** *cinn de bhuaibh*
three heads of cow
'three cows'

Let us now turn to complex numerals. We start with a multiplicative numeral in (11). What the example shows is that regardless of whether the numeral is used in a counting sequence, as in (11a), or as an object-counting numeral, as in (11b), there is no *a* preceding the numeral 'three' or 'hundred.' This means that the numeral *trí* is in the object-counting shape.

(11) Irish (Dylon and Ó Cróinín 1961: 136–137, Stenson 2020: Ch. 20.1)

a. ***trí*** *chéad*
three hundred
'three hundred'

b. ***trí*** *chéad* *bó* *mhór*
three hundred large cows
'three hundred large cows'

For additive numerals, our sources show variation. In Dylon and Ó Cróinín (1961: 136), additive numerals such as 'thirteen' are given bare, without the preceding *a*. However, in Stenson (2020: Ch. 20.1), the counting sequence is given with *a* preceding the complex numeral. This is reflected in (12a) by placing the *a* in parentheses.

(12) Irish (Dylon and Ó Cróinín 1961: 136–137, Stenson 2020: Ch. 20.1)

a. (**a**) **trí** déag
NBR three ten
'thirteen'

b. **trí** bhád déag
three boat ten
'thirteen boats'

However, the crucial data concerns object-counting use, see (12b). The first relevant fact is that the modified noun comes in between the two numerals. This fact alone is hard to reconcile with the idea that the complex numeral forms a constituent to the exclusion of the noun, a point that has been raised for Scottish Gaelic and Biblical Welsh in Ionin and Matushansky (2018: 125). The second relevant fact is that there is not a single NBR *a* in (12b): both numerals are in the object-counting shape.

In sum, Irish is a language where each simple numeral inside the complex is in an object-counting form. Irish is the only language in our sample with this property. This exceptional property of Irish could, perhaps, be linked to another exceptional property, which is that Irish is the only language in our sample where abstract-counting numerals (which are marked by *a* in Irish) are morphologically more complex than object-counting numerals.[2]

2.2. Patterns with abstract-counting numerals

In Bangla, Japanese, Korean, Mandarin, Thai, Telugu and Vietnamese, we find complex numerals that contain at least one abstract-counting numeral. According to the reasoning in

[2] See Wągiel and Caha (2021) for an empirical discussion of the morphological patterns. The authors explain the frequent pattern where object-counting numerals are derived from abstract-counting numerals by proposing that object-counting numerals are semantically derived from abstract-counting numerals by the addition of a dedicated meaning component. The Irish situation apparently goes against this proposal. At the same time, it must be mentioned that current morphological theories contain various ideas how the realisation of a complex structure may contain fewer morphemes than the realisation of a complex structure, see, e.g., Blix (2022) as well as Wągiel and Caha's (2020) treatment of the distinction between *ein* and *eins* 'one' in German.

Section §1, this is only compatible with Analysis A. We present here Mandarin and Vietnamese as two examples of such languages.

Let us begin by showing the distinction between simple abstract- and object-counting numerals in Mandarin (building mainly on He 2015). In (13a), we can see that the abstract-counting numeral for 'two' is *èr*.

(13) Mandarin (Po-Ching and Rimmington 2015: 26; He 2015: 198)
a. *yī jiā yī děngyú* ***èr.***
one add one equals two
'One plus one is two.'
b. ***liǎng gè*** *xuéshēng*
two CLF student
'two students'

The object-counting form of 'two' in (13b) shows two differences. First, the classifier *gè* appears. Second, the numeral has the suppletive form *liǎng*. It is impossible to use *liǎng* in (13a), and it is impossible to use *èr* in (13b).

Let us now turn to complex numerals, starting with multiplicative complex numerals in (14). In (14a), we show an abstract-counting numeral, where *èr* 'two' is the multiplier of 'ten'.

(14) Mandarin (He 2015: 190, 195)
a. ***Èr*** *shí sān shì sùshù.*
two ten three be prime.number
'Twenty three is a prime number.'
b. ***èr*** *shí sān wàn gè xuéshēng*
two ten three 10,000 CLF student
'230,000 students'

The same form (*èr*) also appears in the object-counting numeral in (14b). We can see that the object-counting status of the whole complex numeral is signaled by the classifier that always appears in between the numeral and the counted noun. However, the presence of the classifier does not influence the shape of the numeral 'two', which is internal to the complex numeral and remains in the abstract-counting shape *èr*. Further, there is no classifier in between 'two' (*èr*) and 'ten' (*shí*). This also suggests that 'two' is not counting 'tens' in the same way in which it is counting 'students'. Mandarin multiplicative numerals thus provide (at least) two kinds of evidence in favor of a structure where the numeral first forms a complex constituent consisting of abstract-counting numerals. The two types of evidence are (i) suppletion of 'two' and (ii) the absence of classifiers inside multiplicative numerals.

Finally, let us discuss additive numerals. In (15a), we present the object-counting form of the numeral 'fifty-two'. Interestingly, the numeral 'two' is expressed by *èr*, which is the abstract-counting shape, recall (13a). The object-counting *liǎng* cannot be used here, see (15b).

(15) Mandarin (He 2015: 198)
a. *wǔ shí* ***èr*** *gè xuéshēng*
five ten two CLF student
'fifty-two students'

b. **wǔ shí **liǎng** gè xuéshēng*
five ten two CLF student
Intended: 'fifty-two students'
c. ****èr gè** xuéshēng*
two CLF student
Intended: 'two students'

Importantly, example like (15a) cannot originate from a structure like [fifty students and two students]. If (15a) were to be derived from such a structure, the second member of such a hypothetical coordination would have to be the sequence **èr gè xuéshēng*, but such sequence is ungrammatical, see (15c).

The remaining languages in our sample are classifier languages that show similar distribution of classifiers as Mandarin (but lacking suppletion). We group them together with Mandarin based on the fact that the classifier never appears inside the complex numeral.

As an example of such a language, we provide Vietnamese (Austroasiatic). Vietnamese is a classifier language with only a single classifier for the whole complex numeral. To illustrate the pattern, let us look at bare numerals such as *ba* 'three'. The bare form is used in the counting sequence (16), in arithmetic statements, see (17), and in complex numerals, see (18).

(16) Vietnamese (Ngo 2020: 5)
ba
'three'

(17) Vietnamese (Wągiel and Caha 2021: 477)
Bốn** nhân **hai** bằng **tám
four times two equal.to eight
'Four times two is eight.'

(18) Vietnamese (Ngo 2020: 6, 101)
a. *mười **ba***
ten three
'thirteen'
b. ***ba** mười*
three ten
'thirty'
c. ***ba** trăm*
three hundred
'three hundred'

Bare complex numerals are also used in arithmetic statements such as (19), something that we have also observed for Mandarin, recall (14a).

(19) Vietnamese (Ngo 2020: 6, 101)
*Mười lăm chia cho **ba** **mươi** bằng không phẩy năm.*
ten five divided by three ten equal zero point five
'Fifteen divided by thirty is zero point five.'

When numerals are used as object-counting numerals, a classifier appears after the numeral, see (20a).[3] The example would be ungrammatical without the classifier.

(20) Vietnamese object-counting numerals (Ngo 2020: 21, Trang Phan, p.c.)
a. ***ba*** *bức thư*
three CLF letter
'three letters'
b. ***mười ba*** *bức thư*
ten three CLF letter
'thirteen letters'

Importantly, complex numerals require only one classifier, placed between the whole complex numeral and the noun, see (20b). Wągiel and Caha (2021: 480–482) interpret this to mean that the internal structure of such complex numerals contains abstract-counting numerals. The main indication is the absence of classifiers internally to the complex numeral.[4]

Patterns similar to Vietnamese are found also in Bangla (Biswas 2016: 94–95), Japanese (Wągiel and Caha 2020: 204), Telugu (Lisker 1963: 112), Korean (Lee and Ramsey 2000: 95–96), and Thai (Smyth 2002: 172–173, Wągiel and Caha 2020: 205).[5]

2.3. Non-uniform patterns

Taking the patterns seen in Sections §2.1 and §2.2 at face value, it seems that the structures proposed by Analysis A and B must both be available in individual languages. This conclusion is supported by the fact that sometimes we observe the two different structures even within a single language. We discuss such languages in this section.

As the first case, consider Zulu (Stuart 1940: 41–43). The relevant data set is in (21).

(21) Zulu (Stuart 1940: 41–3, 109)
a. ***ama****-tunga* ***ama****-bili*
class.6-bucket CLASS.6-two
'two buckets'

---

[3] The classifier is obligatory for some nouns, but may be optional with others, see Simpson and Ngo (2018), Phan (2019).

[4] We assume that the numeral and the classifier form a constituent to the exclusion of the noun, following the ideas explored, e.g., in Krifka (1995), Bale and Coon (2014), He (2015), Sudo (2016), Cinque (2021), and Wągiel and Caha (2021). Under this view, the classifier serves as a device that turns the abstract-counting numeral into an object-counting one. Under this view, the fact that we do not find classifiers internally to the complex numeral leads us to conclude that the complex numeral contains abstract-counting numerals.
However, it must be mentioned that there is an alternative analysis of classifiers in the literature, according to which the classifier forms a constituent with the noun (excluding the numeral), as proposed, e.g., in Borer (2005), Chierchia (1998, 2010), Rothstein (2010), Li (2011), Scontras (2013). If that is so, one could hypothesize that the absence of the classifier is due to some other factor. For instance, for additive numerals, one could propose that ellipsis (or right rode raising) eliminates both the noun and the classifier in the first conjunct. As far as we can tell, this approach still cannot handle the data in (15), where the issue is not only the presence/absence of the classifier, but the fact that the numeral 'two' has a dedicated abstract-counting form in the complex numeral.

[5] We are grateful to Priyanka Biswas and Inkie Chung for help with Bangla and Korean, respectively.

b. ***n**-komo* ***m**-bili*
CLASS.10-cow CLASS.10-two
'two cattle'

c. *Ngi funa osheleni **aba** [ishumi] na [m-bili]*
I want shillings CLASS.2 ten and two
'I want twelve shillings. '

d. *Ngi funa [**osheleni aba** ishumi] n- [**osheleni aba** bili]*
I want shillings CLASS.2 ten and- shillings CLASS.2 two
'I want twelve shillings.'

(21a,b) show that Zulu numerals such as 'two' agree with the modified noun in class. In (21a), the numeral has the class 6 prefix, in (21b), it has the class 10 prefix, reflecting the class of the head noun. Different numerals in Zulu differ slightly in what kind of agreement they take depending on their precise numerical value, but we shall skip these details in the interest of space. The important thing is that in additive numerals, Zulu uses two strategies. The first one is illustrated in (21c). It consists in joining the numerals *ishumi* 'ten' and *-bili* 'two' by the conjunction *na-*, which leads to the emergence of a nasal prefix on the numeral *bili* 'two'. This may be interpreted as the emergence of a default class 10 marker, recall (21b), but we shall not dwell on this. Importantly, there is only a single agreement marker *aba* preceding the whole numeral 'twelve', suggesting that it is the numeral as a whole what modifies the noun.

Interestingly, there is an alternative strategy, shown in (21d), with the noun 'shillings' occurring twice, which also leads to the emergence of two agreement markers. We conclude that the existence of two different structures within a single language supports the idea that UG allows for both types of structures. This hypothesis, recall, is independently suggested by the fact that we register different types of languages cross-linguistically.

There are more languages in our sample where complex numerals constitute a non-uniform category, and therefore cannot be easily subsumed either under Analysis A or Analysis B. These are Javanese, Romanian, Maltese and Puyuma. We shall now discuss these in more detail.

Javanese belongs to the Austronesian language family. It is spoken in Java and across the whole Indonesia. We focus on the numeral 'two', which has three different shapes. We give them in (22).

(22) Javanese 'two' (Robson 1992: 75–76)

a. ***ro***

b. *lo-**ro***

c. ***ro**-ng*

The first shape *ro* can only be used in the counting sequence, and we, therefore, classify it as an abstract-counting numeral. The form *loro* in (22b) is ambiguous, and it can be used both as an abstract- and object-counting numeral. Finally, *rong* is only used as an object-counting numeral.

In additive complex numerals, we find the shapes *loro* or *ro*, depending on the numeral, see (23). Each numeral in (23) can be used both as an abstract- or object-counting numeral.

(23) Javanese (Robson 1992: 75–76)
a. ***ro**-las*
two-teen
'twelve'
b. *sawidak* ***loro***
sixty two
'sixty-two'

In the numeral 'twelve' in (23a), we see the shortest form of 'two,' namely *ro*, which can only be used as an abstract-counting numeral, recall (22a). Therefore, it is impossible to analyze *ro-las* as a case where each of the simple numerals combines syntactically with a noun (as in Analysis B), because *ro* never combines with nouns as such (only *loro* does).

The numeral 'sixty-two' in (23b) contains the shape *loro*. This shape is ambiguous between the object- and abstract-counting meaning, and we cannot, therefore, conclude anything here.

In multiplicative numerals, we only find the object-counting shape *rong*, see (24a,b). The shape *rong* is unambiguously object-counting, see (24c).

(24) Javanese (Robson 1992: 75–77)
a. ***rong*** *puluh*
two ten
'twenty'
b. ***rong*** *atus*
two hundred
'two hundred'
c. ***rong*** *kilo*
two kilo
'two kilos'

We thus conclude that multiplicative numerals contain an object-counting shape of the numeral, unlike what we see with the additive numeral in (23a). Javanese is therefore a non-uniform language, where additive numerals conform to Analysis A, while multiplicative structures conform to Analysis B.

We also analyze Romanian in this manner. In (25a), we can see the abstract-counting shapes of the numerals 'one' and 'two'. We can see that for 'one', this form is different from the form used in the object-counting use, see (25b). In the complex numeral 'twenty-one', we see the abstract-counting shape of 'one'. In addition, the counted noun following 'one' is in the plural, and a linker *de* appears; all of these facts are characteristic for higher numerals.

(25) Romanian ((a) Carmen Savu, p.c., (b-c) Ionin and Matushansky 2018: 130–131)
a. ***Unu*** *plus unu fac* ***doi***.
one plus one make two
'One plus one is two.'
b. ***un***/**unu baiat*
one.M boy
'one boy'

c. ***două**-zeci* *și* ***unu*** *de băieți*
two.F ten.PL and one.M of boy.PL
'twenty one boys'

However, the numeral 'two' uses the feminine form before the numeral 'ten' in (25c). This shape is classified by us as an object-counting shape, and we therefore conclude that Romanian is a mixed language: like Javanese, it shows some evidence for abstract-counting forms in additive numerals, and object-counting forms in multiplicative numerals.[6]

Let us now turn to Maltese. Maltese is an Afroasiatic language spoken in Malta. Below, we discuss the pattern of the numeral 'four'. In (26a), we present its abstract-counting shape. In (26b,c), we see two different forms of the object-counting form *erba(t)*. The presence/absence of *-t* apparently relates to the phonological shape of the following noun.

(26) Maltese (Azzopardi-Alexander and Borg 1997: 221, 266, 270; Aquilina 1965: 119, 123)
a. *tnejn* *għal* *tnejn* ***erbgħa***
two times two four
'Two times two is four.'
b. ***erbat*** *irġiel*
four men
'four men'
c. ***erba'*** *soldi*
four pence
'four pence'

Note that the object-counting form in (26b,c) lacks *-għ-*: the presence/absence of this segment helps us identify the type of numeral. Let us now turn to complex numerals. The abstract-counting shape *erbgħa* (truncated due to the loss of the final vowel) can be seen in the multiplicative numeral 'forty' in (27a). We show its object-counting use in (27b).

(27) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)
a. ***erbgħ**-in*
four-ty
'forty'
b. ***erbgħ**-in* *kilo*
four-ty kilo
'forty kilos'

[6] See Ionin and Matushansky (2018: 129–131) for an analysis where the numerals in examples like (25c) are analyzed as 'nominal', rather than abstract-counting numerals. An anonymous reviewer further suggests that the morpheme *un* 'one' may be an indefinite article, rather than a numeral. However, note that *un* 'one' has uses suggesting a numeral status (at least as an option). For instance, it is possible to provide (i) as an answer to the question *How many boys did you see*?

(i) *Am* *văzut* *fix* *un* *băiat*
have.1SG seen exactly one boy
'I saw exactly one boy.'

Such modification would be impossible for an indefinite article (consider the impossibility of **I saw exactly a boy* as an answer to *How many boys did you see?*).

On the other hand, the complex numeral 'fourteen' (28a) contains the object-counting shape *erbat*. The object-counting shape of 'fourteen' adds *-il* to the abstract-counting numeral, see (28b). The important point is that in the object-counting use, both parts of the complex numeral are in the object-counting shape. This can be explained if the structure is as in (28c).

(28) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)
a. ***erbat****-ax*
four-teen
'fourteen' (abstract-counting)
b. ***erbat****-ax-il*
four-teen-il
'fourteen' (object-counting)
c. [erbat N] [ax-il N]

The object-counting form *erba'* is also found in the multiplicative numeral 'four hundred'. We present the abstract-counting shape in (29a) and the object-counting form in (29b). In (29b), *mitt* is the object-counting form of *mija* 'hundred'.

(29) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)
a. ***erba'*** *mija*
four hundred
'four hundred'
b. ***erba'*** *mitt* *liyra*
four hundred pounds
'four hundred pounds sterling'

To summarize, the Maltese numeral 'four' provides us with mixed results: in some multiplicative numerals, recall (27), it is in the abstract-counting form, whereas in other numerals, see (28–29), the object-counting form is found. We found a similar system in Puyuma (Fang-Ching Teng 2007: 108–111).

Summarizing, in Sections §2.1 and §2.2, we saw that some languages use abstract-counting numerals in complex numerals, while other languages use object-counting numerals. In this section, we discussed languages where the two systems coexist. Sometimes they are apparently in free variation (Zulu), whereas in other cases they are fixed for a particular numeral type (Javanese, Romanian, Maltese, Puyuma).

2.4. An unexpected type

Finally, our sample also contains three languages where complex numerals contain component parts that are neither abstract-counting or object-counting numerals. One language like that is Palikur, an Arawakan language spoken in French Guyana and Brazil (Launey 2003). We start our discussion by presenting the object-counting forms of the numeral 'one' in (30).

(30) Palikur (Launey 2003: 121–122)
a. ***paha****-kti* *pilatno*
one-CLF banana
'one banana tree'
b. ***paha****-t* *pilatno*
one-CLF banana
'one banana'

The numeral root *paha* has a suffix that functions as a prototypical classifier and determines the type of object that is quantified over. When it is a 'banana tree', the numeral has the suffix *-kti* (used for arboreal entities). When it is a single banana, the suffix *-t* (for elongated objects and body parts) appears.

More classifiers are illustrated in (31), using the numeral 'two'. Note that with this numeral, classifiers are infixed. In (31a), we see a classifier for clothing and piles, *-rik*. In (31b), there is a classifier for flat objects, *-ka*.

(31) Palikur (Launey 2003: 121–122)
a. ***pi****-rik-****na*** *kagta*
two-CLF-two book
'two piles of books'
b. ***pi****-ka-****na*** *kagta*
two-CLF-two book
'two books'

Let us now turn to multiplicative numerals, see (32a) and (32b).

(32) Palikur (Launey 2001: 38, 45)
a. ***pina*** *madikwa*
two ten
'twenty'
b. ***pima****-vut* *sah*
two-times hundred
'two hundred'

The example in (32a) shows the bare numeral root *pina* 'two' (without any infix) as a multiplier of the numeral 'ten'. The absence of a classifier (which always accompanies object-counting uses) suggests that *pina* is not an object-counting numeral. Similarly, in (32b), the marker *-vut* is translated in the reference grammar as a multiplicative marker analogous to the English 'times'. We do not treat it as a classifier, since it does not infix into the numeral, and it attaches to what appears again as a bare root of the numeral. Therefore, the conclusion is that multiplicative numerals in Palikur do not contain object-counting numerals.

At the same time, they do not contain abstract-counting numerals either. The reason is that when numerals are recited in a sequence, they have a classifier, see (33), contrasting with (32).[7]

[7] Palikur numerals *madikwa* 'ten' and *sah* 'hundred' take no classifiers and show, therefore, no difference between object-counting and abstract-counting forms. Numerals like '(one) hundred two' therefore have no

(33) Palikur (Launey 2003: 114)

a. ***paha**-t*
one-CLF
'one'

b. ***pi**-ta-**na***
two CLF-two
'two'

Summarizing, we conclude that Palikur multiplicative numerals contain neither object- nor abstract-counting numerals.[8]

A similar situation is found in Shuhi (Qi and He 2019). In this language, numerals are always followed by a classifier both when they are used for abstract counting (34a) or for object counting (34b,c). The object-counting numerals take a variety of classifiers, compare (34b,c); for the abstract-counting (34a), the so-called default classifier must be used.

(34) Shuhi (Qi and He 2019: 66–69)

a. **ʤi̱^{35}-ko^{35}**-rɛ^{33} **ʤi̱^{35}-ko^{35}**-ɦõ^{33} mɛ^{33}-ba^{33}-le^{55} ɳɛ^{33}-ko^{35} lɛ^{33}-ʑiŋ^{33}-ʤõ^{33}.
one-CLF-ABL one-CLF-LOC DIR-add-AUX two-CLF DIR-become-DUR
'One plus one is two.'

b. jo?^{33}kuɐ^{55} **ʤi^{33}-pha^{33}**
stone one-CLF
'one lama'

c. sɑ^{55}zə?^{55} **ʤi^{33}-zə?^{55}**
tree one-CLF
'one tree'

As a result, numerals never occur on their own; they are bound roots that always require a classifier or another morpheme. This is similar to Palikur, which also has a classifier both for object and abstract counting, recall (31) and (33). The only place where the classifier is not needed with numeral roots is in complex numerals. In these cases, numeral roots occur bare (35a) and only one classifier for the whole phrase is found (35b).

(35) Shuhi (Qi and He 2019: 66–69)

a. ʤi^{33} ɕe^{55}
one hundred
'one hundred'

b. sɑ^{55}zə?^{55} ɦɑ?^{35}-qe^{33} ɳi^{55} sɐ^{55}-zə?^{55}
tree five-ten CONJ three-CLF
'fifty-three trees'

An additional language that represents this type in our sample is Huehuetla Tepehua (Kung 2007: 480–482).

---

object-counting classifier on 'hundred,' and only have a classifier on 'two.' Thus, the only thing we only know is that the complex numeral contains one object-counting numeral. But since we cannot decide on the form of 'hundred' or 'ten', we cannot assign such complex numerals to any type. The same is true for Swahili (Almasi et al. 2014: 183, 187–188).

[8] There is an alternative form for 'two hundred', the French-creole based form *de-sah.*

3. Conclusion

In the current literature, there are contrasting proposals concerning the structure of complex numerals. Under one analysis, the complex numeral is formed as a constituent first, and it is then used to quantify over objects denoted by the modified noun, see (36a) for an additive complex numeral. The alternative is that each numerical element within the complex numeral combines independently with the modified noun one copy of which is subsequently deleted, see (36b).

(36) Additive complex numerals
a. [20 + 3] cats → 23 cats
b. [20 cats] + [3 cats] → 23 cats

The same two analyses can be advanced also in the domain of multiplicative complex numerals, see (37).

(37) Multiplicative complex numerals
a. [3 × 10] cats → 30 cats
b. 3 × [10 cats] → 30 cats

This paper tried to shed some light on the question of which structures are used in languages by looking at the distinction between abstract-counting numerals, i.e., expressions referring to abstract arithmetical concepts, and object-counting numerals, i.e., forms that modify nouns in order to quantify over entities. Nothing else said, the analyses in (36b) and (37b) predict that complex numerals contain object-counting numerals. On the other hand, (36a) and (37a) predict that complex numerals contain at least one abstract-counting numeral.

Our investigation revealed that the structures in (36a) and (37a) are supported by data from Mandarin, Vietnamese and other classifier languages, where complex numerals contain abstract-counting numerals (§2.2). On the other hand, in Irish, complex numerals generally conform to the templates in (36b) and (37b), recall §2.1.

In yet other languages (we discussed Maltese, Javanese, and Zulu), the different types seem to be mixed within a single language (§2.3). For Shuhi, Palikur and Huehuetla Tepehua, we could not reach any definitive conclusion (§2.4).

Considered in their totality, the patterns found in our sample do not support the idea that languages use only one type of structure. Rather, it seems that both possibilities of forming complex numerals are employed, with languages choosing one or the other structure in ways that we do not fully understand. Quite likely, the investigation of other diagnostics (i.e., other than the abstract/object-counting distinction) may shed more light on this issue.

One thing we would like to emphasize again at the end is that our study revealed that the majority of complex numerals in our sample contain abstract-counting numerals. One could speculate that languages, therefore, prefer to use the abstract-counting structures whenever possible, and only if this strategy is not available for reasons to be understood, the more complex structure is used.

References

Almasi, O., Fallon, M. D., and N. P. Wared. 2014. *Swahili Grammar for Introductory and Intermediate Levels.* Lanham: University Press of America.

Aquilina J. 1965. *Teach Yourself Maltese*. London: The English Universities Press LTD.

Azzopardi-Alexander, M., and A. Borg. 1997. *Maltese.* Abingdon: Routledge.

Bale, A., and J. Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4): 695–707.

Biswas, P. 2016. *Number marking and definiteness in Bangla*. Doctoral dissertation, University of Southern California.

Blix, H. 2022. Interface Legibility and Nominal Classification: A Nanosyntactic Account of Kipsigis Singulatives. *Glossa: A journal of general linguistics* 7(1).

Borer, H. 2005. *Structuring Sense I: In Name Only*. Oxford: Oxford University Press.

Bultinck, B. 2005. *Numerous meanings: The meaning of English cardinals and the legacy of Paul Grice.* Amsterdam: Elsevier.

Chierchia, G. 1998. Plurality of mass nouns and the notion of 'semantic parameter'. In S. Rothstein (ed.), *Events and Grammar,* 53–103. Dordrecht: Kluwer.

Cinque, G. 2021. *On Linearization. Towards a restrictive theory*. Ms., University of Venice.

Dylon, M., and D. Ó Cróinín. 1961. *Teach Yourself Irish*. London: Hodder and Stoughton.

Ionin, T., and O. Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23(4): 315–360.

Ionin T., and O. Matushansky. 2018. *Cardinals: The Syntax and Semantics of Cardinal-Containing Expressions*. Cambridge, MA: MIT Press.

He, C. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164: 189–214.

He, C., and O.-S. Her. to appear. Complex numerals in English: Constituents or not? *English Language and Linguistics*.

Krifka, M. 1995. Common nouns: A contrastive analysis of Chinese and English. In G. N. Carlson, and F. J. Pelletier (eds.), *The Generic Book,* 398–411. Chicago: University of Chicago Press.

Kung, S. S. 2007. *A Descriptive Grammar of Huehuetla Tepehua*. Austin. PhD thesis, University of Texas at Austin.

Launey, M. 2003. *Awna parikwaki. Introduction à la langue palikur de Guyane et de l'Amapa*. Paris: Institut De Recherche Pour Le Développement.

Lee, I., and R. Ramsey. 2000. *The Korean Language.* Albany: State University of New York.

Li, X. 2011. *On the Semantics of Classifiers in Chinese*. Ph. D. thesis, Bar-Ilan University, Ramat Gan.

Lisker, L. 1963. *Introduction to Spoken Telugu*. New York: American Council of Learned Societies.

Ngo, B. 2020. *Vietnamese: An Essential Grammar*. Abingdon: Routledge.

Po-Ching, Y., and D. Rimmington. 2015. *Chinese: A comprehensive grammar*. London and New York, NY: Routledge.

Phan, T. 2019. The absence of classifiers in numeral classifier constructions in Vietnamese. In M. Bağrıaçık, A. Breitbarth, and K De Clercq (eds.), *Mapping Linguistic Data,* 208–213. Online publication (https://biblio.ugent.be/publication/8625919/file/8625920.pdf).

Qi, J., and C. He. 2019. The morphosyntax of numerals dz̄i33/dz̄i35 'one' in Shuhi and implications for the semantics of numerals. *Lingua* 225: 63–80.

Robson, S. 1992. *Javanese Grammar for Students*. Melbourne: Monash University.

Rothstein, S. 2010. Counting and the mass/count distinction. *Journal of Semantics* 27(3): 343–397.

Rothstein, S. 2017. *Semantics for counting and measuring*. Cambridge: Cambridge University Press.

Scontras, G. 2013. Accounting for counting: A unified semantics for measure terms and classifiers. In T. Snider (ed.), *Proceedings of Semantics and Linguistic Theory 23,* 549–569. Ithaca, NY: CLC Publications.

Simpson, A., and B. Ngo. 2018. Classifier syntax in Vietnamese. *Journal of East Asian Linguistics* 27(3): 211–246.

Smyth, D. 2002. *Thai. An Essential Grammar*. London: Routledge.

Stenson, N. 2019. *Modern Irish: A comprehensive grammar*. Oxon: Routledge.

Stuart, P. A. 1940. *A Zulu Grammar for Begginers.* Pietermaritzburg: Shuter & Shooter.

Sudo, Y. 2016. The semantic role of classifiers in Japanese. *Baltic International Yearbook of Cognition, Logic and Communication* 11(1): 1–15.

Tatsumi, Y. 2021. *Linguistic Realization of Measuring and Counting in the Nominal Domain: A Cross-linguistic Study of Syntactic and Semantic Variations*. PhD dissertation, University of Connecticut.

Wągiel, M., and P. Caha. 2020. Universal semantic features and the typology of cardinal numerals. *Catalan Journal of Linguistics* 19: 199–229.

Wągiel, M., and P. Caha. 2021. Complex simplex numerals. *Acta Linguistica Academica 68*(4): 470–515.