Table of Contents

Paradigmatic Morphology
Splinters, Combining Forms, and Secreted Affixes

Elisa Mattiello, University of Pisa

Splinters, combining forms, and secreted affixes are three morpheme (or morpheme-like) elements which are often conflated in the literature on English word-formation. Scholars have differently focused on their morphological origin (i.e. blending, paradigmatic substitution, analogy) or on their semantics (i.e. secretion vs. mere abbreviation) (Warren 1990; Fradin 2000; Mattiello 2007; Bauer et al. 2013).
This paper investigates these phenomena as part of paradigmatic morphology, or similarity among words. In particular, the investigation of five case studies (i.e. -(a)holic, docu-, -exit, -umentary, -zilla) shows that they are frequently used to create new words and even to produce series, through analogy via schema (cf. Köpcke 1993, 1998). In the paper, diachronic study combined with corpus-based analysis help us 1) categorise these phenomena as 'marginal' vs. 'extra-grammatical', and as 'productive' vs. 'creative', and 2) shed some light on their role in the development of morphological rules and in the expansion of the English lexicon.

Keywords: *Paradigmatic morphology, Splinters, Combining forms, Secreted affixes, Blending*

1. Introduction

This study deals with the part of morphology which is referred to as paradigmatic morphology, pioneered by van Marle (1985) and "based on some sort of resonance or similarity between words in the lexicon" (Bauer et al. 2013: 519). Phenomena of paradigmatic morphology are generally classified as "word creation" (Ronneberger-Sibold 2000, 2008) or "extra-grammatical morphology" (Dressler 2000; Mattiello 2013), as opposed to regular (i.e. productive) English word-formation (Bauer 1983; Plag 2003; Bauer et al. 2013).

The theoretical framework adopted in this study for the analysis of such phenomena is Natural Morphology (Dressler et al. 1987; Kilani-Schoch 1988; Dressler 2000; Kilani-Schoch & Dressler 2005). Within this framework, prototypical grammatical morphology is distinguished from both extra-grammatical and marginal morphology (Doleschal & Thornton 2000). Extra-grammatical morphology applies to a set of heterogeneous formations (of an analogical or rule-like nature) which do not belong to morphological grammar, in that the processes through which they are obtained are not clearly identifiable and their input does not allow a prediction of a regular output like rules do (Mattiello 2013: 1). For instance, the blends *brinner* ← *br*(eakfast + d)*inner* [2008] 'breakfast eaten at dinner time' (*Urban Dictionary*) and *blaxploitation* ← *bla*(ck + e)*xploitation* [1972] 'the exploitation of black people' (OED2) belong to extra-grammatical morphology, in that they are only partially predictable, whereas the regular derived words *mini-break* or *breaker* or the compound *blackbird* are fully predictable from their inputs. Within the same framework, marginal (but still grammatical) morphology applies to phenomena which are non-prototypical (i.e. at the boundaries) of morphology (Dressler 2000: 6–7), in that they are transitional between morphology and other linguistic levels (e.g. lexicon, syntax) or between the subcomponents of morphology (i.e. inflection, derivation, and compounding). For instance, the new words *read-o-holic* [2013]

(*The Guardian*) and *pizza-holic* [2015] (CNN) have been coined after the model of *workaholic* [1947], *sugarholic* [1955], *foodaholic* [1965], etc., all exhibiting the "final combining form" -(*a*)*holic* (← *alcoholic*, Warren 1990; called "suffix" in the OED). Specifically, combining forms are transitional between derivation and compounding, depending on whether we consider -(*a*)*holic* to be a bound or a free morpheme.

In particular, the study explores the creative formation of new words by means of the blending process. Blending is generally regarded as a "creative technique" (Ronneberger-Sibold 2008) used to produce new lexemes in domains such as humorous literary texts and brand names (Kemmer 2003; Lehrer 2003, 2007; Gries 2004, 2012). However, recent studies show that blends exhibit (sub)regularities and tendencies, especially in terms of prototypical patterns and phonological regularity (Mattiello 2013), prosodic structure (Arndt-Lappe & Plag 2013), but also of frequently occurring "splinters" (Lehrer 1996, 2007) or "secreted affixes" (Fradin 2000). Instances of splinters include -*arian* (← *vegetarian*), as in *fruitarian* [1893], *nutarian* [1909], etc. (OED2-3), *docu-* (← *documentary*), as in *docudrama* [1961], *docusoap* [1979], etc. (OED2–3), and -*exit*. The latter, despite its existence as an independent word, can be reinterpreted as an affix, especially a secreted one, whose meaning is not simply 'leave', but 'withdrawal from the European Union'. This meaning is illustrated both by English neologisms, such as *Grexit* ← *Greece*/*Greek* + *exit*, *Brexit* ← (*Great*) *Britain*/*British* + *exit*,[1] both dated [2012] in *Wordspy* and only included in the OED since March 2017, and by occasionalisms, such as the analogical *Spexit* or *Frexit* [2015] 'exit of Spain/France from the EU' recently found in *The Guardian*. The initial splinter counterpart is *Br-* (← *British*), as in *Bremain* or *Brentry* [2016] 'British remain/entry' (*The Guardian*), both obtained analogically after the model of *Brexit*. Analogy, therefore, is the underlying process of these new words, which suggest an analysis in terms of paradigmatic substitution (Bauer et al. 2013; Mattiello 2017).

This study deals with the creation of new words by analogy via schema (see "schema" in Köpcke 1993, 1998), with a set of prototypical words as model. In particular, the study concentrates on three very close products of analogy via schema – namely, splinters, combining forms, and secreted affixes – which are generally confused or partially overlap in the literature on word-formation, with morphologists using different labels to refer to equivalent phenomena, or focusing on different aspects of the same phenomenon, but disagreeing with others on various issues. By investigating these types of new morpheme or morpheme-like element, the study aims at showing that:

- Like productivity (Plag 1999; Bauer 2001), creativity is a scalar concept, ranging from low creativity, as when analogy combines with productive rules (see "creative compounds" in Benczes 2006; Franceschi 2013; Crawford Camiciottoli 2015), to high creativity, as in blending, where no rule applies.
- Analogy is not a strictly local phenomenon, but can give birth to productive series (cf. Bauer 1983: 96), as when splinters, combining forms, or secreted affixes become recurrent in the creation of new words.
- Hence, unlike ad hoc word-formation (see "Ad-hoc-Wortbildung" in Hohenhaus 1996), splinters may trigger a schema model, which is not as abstract as rules, but may represent the first step towards the development of a rule.

---

[1] See Milizia (2014) for the original portmanteau word *Brixit*, which later developed into *Brexit* probably because of the increased similarity with *Grexit*.

Some research questions which will be addressed in this study include:

1) What is the difference between splinter and combining form or secreted affix?
2) Assuming that (Lehrer 2007: 121) is correct that "the transition from splinter to independent morphemehood is a diachronic process", can intermediate stages be identified between novel splinter and fully productive, transparent morpheme?
3) When do we have a shift from a unique model to a set of words as model? (cf. "local" vs. "extended analogy" in Klégr & Čermák 2010; "surface analogy" vs. "analogy via schema" in Mattiello 2017)
4) Is analogy via schema the first phase of the development of a rule?

Corpus-linguistic quantitative analyses in English corpora will provide evidence of the frequency and productivity of some novel splinters, combining forms, and secreted affixes in English. Lexicographic and diachronic study will help shed some light on their origin and categorisation, and, more generally, on the development and expansion of the English lexicon (see Miller's 2014 "lexicogenesis"). In other words, corpus-based analyses should help us discriminate between new words formed by surface analogy, with one precise word that acts as model, from those based on a schema, with several concrete words as model. Moreover, among cases of analogy via schema, or series, quantitative analyses should also clarify the difference between novel splinters and more established combining forms or even suffixes. We hypothesise that the shift from novel splinter to new productive morpheme is a gradual diachronic process, with intermediate stages involving semantic reinterpretation and generalisation.

2. The theoretical landscape

English word-formation has been studied by many scholars (Adams 1973; Bauer 1983; Štekauer 2000; Plag 2003; Štekauer & Lieber 2005; Bauer et al. 2013), whose attention was primarily focused on morphological grammar. Another field of study in morphological research pertains to word creation, which "refers to all operations for the production of new lexemes which are not covered by regular word formation" (Ronneberger-Sibold 2008: 201), including "creative techniques" such as shortening or blending. For these and related creative techniques, Baldi & Dawar (2000: 963) have used the label "unconventional word-formation", in that blending and abbreviations are out of ordinary (conventional) norms and rules, whereas Zwicky & Pullum (1987) classify them as "expressive (vs. plain) morphology", on account of their pragmatic effect. According to Hohenhaus (1996), blending and acronyms are part of what he calls "Ad-hoc-Wortbildung" (ad hoc word-formation), that is, the creative formation of new words, often by means of direct analogy. Aronoff (1976: 20) groups together clippings, blends, and acronyms under the label "oddities", while Bauer (1983: 232) calls them "unpredictable formations", though emphasising that they are so common in English that "it is misleading to consider them out of the ordinary". The present study explores the relationship between blending and the analogical process, showing how the latter affects the former, conferring regularity on blends and increasing their degree of predictability.

In a volume on lexical creativity, Lehrer (2007: 116) defines blends as "underlying compounds which are composed of one word and part of another, or parts of two (and

occasionally three) other words", adding that each word part in a blend is called "splinter". In this study, the term 'splinter' will be likewise used as synonymous with 'blend's part' (cf. its broader sense in Bauer et al. 2013).

Blends have been variously studied for their contribution to neology (Kemmer 2003; Lehrer 2003), for their prosodic structure (Gries 2004, 2012), and their preferred contexts of use, namely literary texts and product names. For instance, in Richard Lederer's book *Adventures of a Verbivore®*, the latter is a blend that merges *verb* and (h)*erbivore*, and, similarly, the creative name of the fruit-flavoured drink *Fruitopia®* originates by blending *fruit* and (u)*topia* (both in Lehrer 2007: 129–130).

While blends are similar to compounds in that they merge two or more words into one, their irregularity and only partial predictability make them out of ordinary English word-formation rules, i.e. extra-grammatical. Hence, blends exhibit a higher degree of creativity than creative (analogical) compounds. Indeed, creativity is a scalar concept (cf. Plag 1999; Bauer 2001; see Ladányi's 2000 productivity-creativity scale) and can involve analogy. For instance, Benczes (2006: 6) labels "creative compounds" those compounds which are based on metaphorical or metonymic associations, such as *scarlet-collar* (worker) [2000] 'a woman who operates an Internet pornographic site' (*Wordspy*), inspired by *white-collar* (worker), *blue-collar* (worker) and similar complex words. However, the formation of the compound adjective *scarlet-collar*, based on the compound family X-*collar* (Mattiello & Dressler forth.), is more predictable and less creative than the formation of a blend such as *boatel* ← *boa*(t + ho)*tel* [1950] 'a boat which functions as a hotel' (OED3), after the unique model *motel* ← *mo*(tor + ho)*tel*.

Nonetheless, recent approaches to blends have conferred some regularity on blends, by claiming that morphological and prosodic factors can influence their structure. For instance, Mattiello (2013) stresses that the prototypical blending pattern retains the initial part of one source word and the end of another, as in *smog* ← *sm*(oke + f)*og* [1905] (OED2). Moreover, Arndt-Lappe & Plag (2013) have proved that 1) blend's length most often conforms to the length of the longer source word, and 2) there seems to be a preference for blends to have no more than three syllables. In this paper, we assume that the regularity and partial predictability of blends are also connected with some frequently occurring splinters, such as *-ercise* ← (ex)*ercise*, found in *sexercise* [1942] 'sexual activity regarded as exercise' (OED3), *dancercise* [1967] 'dancing performed as an exercise' (OED2), and *boxercise* [1985] 'a form of aerobic fitness routine incorporating exercises from boxing' (OED3). As Lehrer (2007: 121) observes: "[w]hen a splinter becomes so common that people start using it frequently, it may lose its connection with the source word and can be considered a morpheme in its own right".

Nowadays some of these splinters have become so common and productive in English that they deserve the label of "combining forms" (Warren 1990) or "secreted affixes" (Fradin 2000). This regularity in use is connected with the analogical process.

2.1 *Blending and analogy*

Although blends should not be confused with, nor conflated with the analogical process (cf. Bauer 1983), some blends are created by analogy or similarity with others. Similarity may be with a unique word: for instance, the above-mentioned *brinner* and *blaxploitation* (see § 1) respectively share the beginning and the end with their models *brunch* ← *br*(eakfast + l)*unch* and *sexploitation* ← *sex* + *exploitation*. Alternatively, analogical words may belong to a series sharing the same formation: e.g., *Frexit* ← *Fr*(ance) + *exit*, *Germexit* ← *Germ*(any) + *exit*, and

Spexit ← *Sp*(ain) + *exit* are all based on *Grexit* and *Brexit* (*The Guardian*, 2015), with the shared portion *-exit* acquiring a specific meaning 'exit from the EU'. With a different shared portion *Br-*, we find, besides the above-mentioned *Bremain* and *Brentry*, the recent *Bremorse* and *Bregret* 'a sense of remorse/regret for leaving the UK' (*The Independent*, 2017). Related humorous wordplays are *BrexPitt* 'the end of Brad Pitt and Angelina Jolie's marriage' and *Bakexit* 'the BBC's loss of The Great British Bake Off' (*The Guardian*, 2016).

The two types of similarity relationships that we find in the above blends can be accommodated within the model of analogy elaborated in Mattiello (2016) and refined in Mattiello (2017) for word-formation. Specifically, Mattiello (2017) distinguishes between:

- Surface analogy (after Motsch's 1981: 101 "Oberflächenanalogie"; cf. "local analogy" in Klégr & Čermák 2010: 235): i.e. the word-formation process whereby a new word is coined that is clearly modelled on an actual model word (e.g. *brinner* after *brunch*);
- Analogy via schema (see Köpcke 1993, 1998 for "schema" in inflectional morphology; "extended analogy" in Klégr & Čermák 2010: 235): i.e. providing a pattern for a series of formations (e.g. *-ercise*, *-exit*).

The focus in the present study is especially on analogy via schema. This type of analogy, based on concrete prototype words such as the *-exit* series, therefore differs from surface analogy, with a unique model (i.e. *brunch* is the only model for *brinner*), and both differ from rules, based on abstract templates. In other words, rules' templates are abstract models, whereas a schema is a concrete model identifiable as two or (preferably) more words. These words may constitute:

- A word family (Mattiello & Dressler forth.): i.e. a group of words sharing some of the same base(s) (e.g. *white-collar* [1911], *blue-collar* [1929], *pink-collar* [1975], *green-collar* [1992], OED3; Benczes 2006: 144–145);
- A series of words sharing the same formation (e.g. *-tainment* ← (enter)*tainment* in *docutainment* [1978] OED2, *infotainment* [1980] OED3, *edutainment* [1983] OED2, *advertainment* [2004] COCA).

2.2 *Series: The literature on splinters, combining forms, and secreted affixes*

In the literature, splinters, combining forms, and secreted affixes are all connected – in different degrees – with the blending phenomenon, with scholars who have shown their interest in these mechanisms since the 1990s.

According to Bauer et al. (2013: 519), splinters belong to paradigmatic morphology, in that they are used to form new words which have some sort of resonance or similarity with other words in the lexicon. They define splinters as "originally (mostly) non-morphemic portions of a word that have been split off and used in the formation of new words with a specific new meaning" (Bauer et al. 2013: 525). For instance, *-gate* ← (Water)*gate* was used with the meaning 'an actual or alleged scandal' (OED2) in words such as *Dallasgate* [1975], *Billygate* [1980], or *Monicagate* [1998], the latter included by Miller (2014: 89) in 'puns'. According to Mattiello (2017), the process that occurs in this word-formation type is a "paradigmatic substitution". In other words, *Monicagate* originated from the substitution of a

first name in the analogical proportion *Billy* (Carter): *Billygate = Monica* (Lewinsky) : X (X = *Monicagate*).[2]

Lehrer (2007: 116) also observes that, "[a]lthough a splinter is a clipping, it cannot occur alone, as a word", but this statement is contradicted by Bauer et al. (2013: 528), who claim that "when [splinters] do become more productive, they may even start a life as a free form". For instance, originally *-burger*, as in *cheeseburger* [1930], *chickenburger* [1936], and *beefburger* [1940], was a splinter coming from the reinterpretation of *Hamburger* (from the German city of *Hamburg*) as *ham + burger*, although there was neither semantic nor morphological connection with *ham*. Later, *burger* became an independent morpheme with the meaning 'patty served on a bun' (Bauer et al. 2013: 528), or simply used as 'a familiar shortening of hamburger', as in this quote from *The Observer* [1960]: "Recently the Hamburger has become just a '*burger*', and there are 'beefburgers', 'chefburgers', 'cheeseburgers', 'eggburgers' and even 'kingburgers'" (OED2).

Bauer et al. (2013) also claim that some splinters can be free forms, such as *exit* in *Brexit*, or *ware*. The latter commonly refers to 'articles of merchandise' in compounds (e.g. *glassware*, *tableware*), but acquires a distinct meaning 'software' when it is used in derived words (e.g. *courseware* [1978] 'computer programs designed for use in an educational course' OED2, *freeware* [1981], *shareware* [1983] 'software which is available free of charge' OED2–3). In the latter case, a compound analysis has to be excluded and a blend analysis *course/free/share* + (soft)*ware* is to be preferred.

Combining forms likewise belong to paradigmatic morphology, in that their origin is analogical. Warren (1990: 115) defines combining forms either as "elements which represent parts of other words" (e.g. *eco*-(logical) in *eco-damage*, (alco)-*holic* in *chocoholic*) or as "elements which from a purely formal point of view are not new morphemes, but which have novel meanings" (e.g. *-gate* 'political scandal').[3] According to this distinction, combining forms can be either abbreviated or secreted: the former are shorter substitutes for their longer counterparts in the combination, while the latter also involve a semantic specialisation. Thus, *eco-* is abbreviated, in that the meaning of *ecological* is entirely retained in *eco-damage*, whereas *-(a)holic* is secreted, in that, e.g. in *chocoholic*, only the semantic elements 'person addicted to' are kept from the meaning of *alcoholic*, but the semantic element 'alcohol' is not (Mattiello 2017: 41). Warren (1990) illustrates this distinction by using the examples of *cheeseburger* (abbreviation) and *fishburger* (secretion), both obtained from the combining form *-burger*. Indeed, while a *cheeseburger* is 'a hamburger with a slice of melted cheese' and could be analysed as *cheese + hamburger*, a *fishburger* is not 'a hamburger with fish', but 'a fried patty made of fish served in a bun'. This testifies that we are not facing a case of abbreviation from *fish + hamburger* (see Fradin 2000: 19–20 for the representations of these words). Therefore, the same combining form can accept either a blend analysis (when abbreviation occurs) or an analysis in terms of paradigmatic substitution and semantic specialisation (when secretion occurs).

Secreted affixes is another label used in the literature to define phenomena of paradigmatic morphology. Fradin (2000: 46) defines secreted affixes (or affix-like forms) as "forms where secretion takes place" and whose "semantics can always be traced back to the meaning of a model-lexeme": e.g., *-speak* ← Orwell's (New)*speak* used for 'a characteristic

---

[2] A similar proportion could be envisaged with other *-gate* formations whose first element is the proper name, nickname, etc., of a person or organisation implicated in the scandal, such as *Cartergate*, *Floodgate*, *Stalkergate*, *Totegate*, etc. These concrete prototype words represent a schema model.

[3] Cf. neoclassical combining forms, such as *bio-* or *-logy*, from Latin or Greek, which are out of interest here.

mode of speaking', as in *computer-speak* [1968] 'computerese' (OED3), *techspeak* [1992] 'technical jargon' (OED3, s.v. *tech*), etc. Fradin (2000) also specifies that secreted affixes differ from blends in several respects. First, semantically, secreted affixes often involve the partial loss of meaning, while the semantic content is kept intact in blends. Second, unlike blends, they always involve abstraction. Third, phonologically, secreted affixes are uniformly obtained by shortening the beginning or the end of a model-lexeme (cf. the various blending patterns in Mattiello 2013).

Remarkably, Fradin (2000) includes among secreted affixes the same elements that are elsewhere described as combining forms (i.e. *-gate*, *-holic*, *-burger*) (e.g. by Warren 1990). However, the overlap is only with secreted combining forms, not with abbreviated ones. Thus, *Euro-* ← *Euro*(pean) in *Euromarket* [1953] (OED3), *nega-* ← *nega*(tive) in *negademand* [1973] (OED3), and *e-* ← *e*(lectronic) in *e-text* [1990] (OED3) are abbreviated combining forms which do not involve a secretion process. By contrast, *-gram* ← (tele)*gram* is a secreted affix (or a secreted combining form) which denotes 'a message delivered by a representative of a commercial greetings company', as in *kissogram* [1982] 'a greetings message sent through a commercial agency, which is delivered with a kiss' (OED2), and similar occasionalisms recorded in OED2: i.e. *Gorillagram*, *Rambogram*, and *strippergram*. An anomalous case is in *-exit*, which does not involve any abbreviation, but only secretion, i.e. the semantic specification 'withdrawal from the EU', and is used as an affix with this meaning in *Grexit*, *Brexit*, etc.

Admittedly, although splinters, combining forms, and secreted affixes can be grouped together as part of paradigmatic morphology, their categorisation and theoretical framing still deserve attention and detailed examination. For instance, one may wonder whether they belong to grammatical, extra-grammatical, or marginal morphology (the latter two defined in 1). The label 'splinter' was originally used only for word parts (Lehrer 1996), so it alludes to the extra-grammatical process of abbreviation involved in blending, whereas 'combining form' and 'affix' respectively refer to the word-formation processes of composition and affixation, generally regarded as grammatical. However, since they are placed at the boundaries between compounding and derivation, combining forms are rather viewed as part of marginal morphology (Dressler 2000; Mattiello 2013, 2017). Finally, secreted affixes stay, according to Fradin (2000: 54), outside morphology, in that they are extra-grammatical means of forming lexemes, although they actually involve a certain level of abstraction and regularity. These dissimilar or even divergent remarks confirm that the three phenomena under investigation here represent an area replete with complexities. The analysis carried out in section 4 is meant to bring more clarity to this area. The methodology for the analysis is explained in section 3.

3. Method

The new words analysed in this study include both neologisms and nonce words or occasionalisms (see Christofidou's 1994 "Okkasionalismen"). The former are new words meant to enrich the lexical stock of a language (Dressler 1993: 5028), whereas the latter are only used on one specific occasion, but are unlikely to become a permanent part of the vocabulary (see also Algeo 1991: 3; Bauer 2001: 39). The focus in this study is on the new words of present-day English, from the second half of the twentieth century to the current (twenty-first) century.

Initially, data selection was made manually, by collecting new words which exhibited a shared segment with others and which were potential candidates for an analysis in terms of paradigmatic substitution and analogy. Sources for data were primarily:

1) *The Oxford English Dictionary Online*, in its continuous update from the second (OED2) to the third edition (OED3);
2) Previous studies or paper dictionaries on the topic of blends and English neologisms (e.g. Algeo 1991; Green 1991; Lehrer 1996, 2003, 2007; Baldi & Dawar 2000; Kemmer 2003; Gries 2004, 2012; Bauer et al. 2013; Mattiello 2013, 2017; Miller 2014);
3) Existing websites on the new words that either have recently entered the English vocabulary or are on their way to, such as McFedries's *Wordspy.com*, or Peckham's *Urban Dictionary*.

For the corpus-based analysis, five case studies (i.e. *-holic*, *-zilla*, *docu-*, *-umentary*, and *-exit*) were selected and investigated in this order. The selection was based on their different classification in the OED, either as more recognised morphological forms (i.e. suffix, combining form) or as more volatile and less predictable ones (splinter or blend's part). Hence, we expected that corpus-based analyses and word frequencies could better clarify these terminological and morphological distinctions.

For the quantitative analysis, data collection and frequency investigation were machine-driven. Automatic search was carried out in two corpora of English, both retrievable from the Brigham Young University's website, namely:

1) *Corpus of Contemporary American English* (henceforth COCA), containing more than 520 million words (20 million words each year 1990–2015) and equally divided among spoken, fiction, popular magazines, newspapers, and academic texts;
2) *News on the Web Corpus* (henceforth NOW), containing 4.4 billion words of data from web-based newspapers and magazines from 2010 to the present time, and growing by about 5–6 million words of data each day (last accessed May 2017).

Each corpus allowed a search for a word or a word part. For the corpus frequency of splinters, combining forms, and secreted affixes, selection was made using the asterisk (*), either preceding or following the word part. Given the different size of the two corpora used for corpus-based analyses, token frequencies were also normalised, either per million words (pmw, COCA) or per billion words (pbw, NOW).

4. Analysis

The analysis conducted in the following subsections is both lexicographic and corpus-based and it is meant to aid categorise the elements selected (i.e. *-holic*, *-zilla*, *docu-*, *-umentary*, and *-exit*) either as productive morphemes with a well-recognised status or as creative non-morphemic elements which are part of the extra-grammatical processes of the language. The rationale behind the order of the case studies analysed is connected to the labels defining each of these elements in the OED, namely "suffix", "combining form", or "blend's part".

4.1 *Case study 1: -(A)HOLIC*

The first case study analysed in the section is *-(a)holic*, from *alcoholic* (with respelling, but also in the variant *-oholic*, with the same pronunciation), which is described in OED3 as a "suffix" "forming nouns (often humorous nonce-words) denoting a person who appears to be addicted to the thing, activity, etc., expressed by the first element". The following examples recorded in the OED are given in chronological order:

- *workaholic* [1947] 'a person addicted to working', *tobaccoholic* [1954], *milka-holic* [1955], *sugarholic* [1955], *cake-aholic* [1957], *chocoholic* [1961] 'a person who is addicted to or very fond of chocolate', *foodaholic* [1965] 'a compulsive eater', *golfaholic* [1971], *carboholic* [1973], *computerholic* [1977], *shopaholic* [1977] 'a compulsive shopper', *newsaholic* [1979], *spendaholic* [1982], *cruise-oholic* [1989], *sexaholic* [1994], and *rage-aholic* [2008] (OED).

This list shows that *workaholic* is probably the antecedent of all others and that *-(a)holic* has viewed its greater expansion between the 1950s and the 1970s. The OED also records a quote where *-(a)holic* is used as a converted noun – "Everybody is an '*olic*' of some kind, alcoholic, *sexaholic*, *workaholic*, *support-groupaholic*" (OED3, 1994), therefore illustrating its use also as an independent morpheme.

Corpus investigation appears to confirm these claims. In COCA, *-(a)holic* is frequently found in both well-established words (e.g. *workaholic*, *chocoholic*) and occasionalisms (e.g. *family-oholic*, *fatheraholic*, *fruitoholic*). Its use as a free form (*aholic*) is also recorded twice. A **holic* search in COCA gives the following results, arranged in order of frequency. Both token frequency and normalised (pmw) frequency are provided in round brackets. Different spellings – i.e. hyphenated or not – account for the same word.

- WORKAHOLIC (444/0.85 pmw), SHOPAHOLIC/SHOP-AHOLIC (62/0.11 pmw), CHOCOHOLIC/CHOCKAHOLIC (31/0.05 pmw), FOODAHOLIC/FOOD-AHOLIC (7/0.013 pmw), RAGE-AHOLIC/RAGEAHOLIC (5/0.009 pmw), SEXAHOLIC (5/0.009 pmw), SPENDAHOLIC (5/0.009 pmw), AHOLIC (2/0.003 pmw), BUYAHOLIC (2/0.003 pmw), GAMBLEAHOLIC/GAMBLAHOLIC (2/0.003 pmw), PLANTAHOLIC (2/0.003 pmw), SPORTSAHOLIC (2/0.003 pmw), DEALAHOLIC (1/0.001 pmw), DOREEN-AHOLIC 'person addicted to Doreen's pizza restaurant' (1/0.001 pmw), ELKOHOLIC (1/0.001 pmw), EVENTAHOLIC (1/0.001 pmw), FABRIC-HOLIC (1/0.001 pmw), FAMILY-OHOLIC (1/0.001 pmw), FATHERAHOLIC (1/0.001 pmw), FISHAHOLIC (1/0.001 pmw), FRUITOHOLIC (1/0.001 pmw), FUNDRAISE-AHOLIC (1/0.001 pmw), GOLFAHOLIC (1/0.001 pmw), GROWTHAHOLIC (1/0.001 pmw), HERBAHOLIC 'person addicted to spices' (1/0.001 pmw), HOARDAHOLIC (1/0.001 pmw), HOOPAHOLIC (1/0.001 pmw).

An additional search in NOW provides comparable results in terms of frequency and productivity. Normalised frequencies are given per billion words (pbw).

- WORKAHOLIC/WORK-A-HOLIC/WORKOHOLIC (2,273/0.59 pbw), SHOPAHOLIC/SHOPOHOLIC (479/0.12 pbw), CHOCOHOLIC/CHOCAHOLIC (199/0.05 pbw), SHAREAHOLIC (101/0.02 pbw), RAPPERHOLIC (65/0.01 pbw),

ECOHOLIC (19/0.005 pbw), RAGEAHOLIC/RAGE-AHOLIC (18/0.004 pbw), AQUAHOLIC (16/0.004 pbw), BOOKAHOLIC (16/0.004 pbw), SEXAHOLIC (16/0.004 pbw), MILKAHOLIC (13/0.003 pbw), SPENDAHOLIC (13/0.003 pbw), APPSAHOLIC (11/0.002 pbw), CRAFTHOLIC 'name of a company producing toys' (9/0.002 pbw), APPLEHOLIC (8/0.002 pbw), FOODAHOLIC (8/0.002 pbw), INFOHOLIC (8/0.002 pbw), SHOEAHOLIC (8/0.002 pbw), TWITTERHOLIC (8/0.002 pbw), COFFEEHOLIC (7/0.001 pbw), PLANTAHOLIC (7/0.001 pbw), SHOCKAHOLIC (7/0.001 pbw), TWEETAHOLIC (7/0.001 pbw), NEGAHOLIC (6/0.001 pbw), LOVE-AHOLIC (5/0.001 pbw), ROCKAHOLIC (5/0.001 pbw), SEEDAHOLIC 'name of a company's website selling seeds' (5/0.001 pbw), SUGARHOLIC (5/0.001 pbw).

This data shows that -(*a*)*holic* is not only a recognised element recorded in eminent dictionaries such as the OED, but also a morpheme attested in corpora and whose profitability is undeniable. The meaning associated to this morpheme – 'person addicted to' – entails a generalisation semantic process which accounts for its high level of abstraction. The bases to which -(*a*)*holic* is added are commonly nouns (*apple*, *coffee*, *plant*), even abbreviated ones (*apps*, *info*), rarely, personal names (*Doreen*), verbs (*spend*), or adjectives (*nega*(tive attitude) 'pessimism'). The pattern used to obtain new words, even occasionalisms, is regular and still novel words are predictable, both in form and in meaning.

4.2 *Case study 2:* -ZILLA

The second case study is -*zilla*, from *Godzilla* (an alteration of the Japanese film *Gojira* 'gorilla', reinterpreted as *God* + *zilla*), which is described in OED3 as a "combining form" "forming humorous, usually temporary words which depict a person or thing as a particularly imposing, relentless, or overbearing example of its kind". The examples recorded in the OED mostly include occasionalisms:

- *Hogzilla* [1978], *Bosszilla* [1988], *Bird-zilla* (referred to a turkey) [1993], *Bridezilla* [1995] 'a woman thought to have become intolerably obsessive or overbearing in planning the details of her wedding', *groomzilla* [2003], *mom-zilla* [2005], *thespzilla* [2007] (OED).

In the following quote, -*zilla* is added to a clipped word from *thesp*(ian), originally theatrical: "That was very Hollywood, however, and this is very British, especially those depressing streetscapes of north London, through which our ageing *thespzilla* stomps defiantly around in his old-geezer's woolly hat" (OED3, 2007).

Corpus investigation in COCA and NOW gives both neologisms (e.g. *bridezilla*, *momzilla*) and occasionalisms (e.g. *teenzilla*) as findings. Results from COCA are:

- BRIDEZILLA (32/0.06 pmw), DOGZILLA (11/0.02 pmw), COWZILLA (7/0.01 pmw), HOLLYZILLA 'name of a monster comparable to Godzilla' (2/0.003 pmw), MOMZILLA (2/0.003 pmw), NUNZILLA (2/0.003 pmw), GROOMZILLA (1/0.001 pmw), HOGZILLA (1/0.001 pmw), TEENZILLA (1/0.001 pmw).

Results from NOW include:

- BRIDEZILLA/BRIDE-ZILLA (269/0.07 pbw), SNOWZILLA 'name of a storm in the US' (76/0.02 pbw), GROOMZILLA (29/0.007 pbw), HOMERZILLA 'name of a doughnut-eating sea monster' (8/0.002 pbw), BATZILLA 'name of a rescue group saving bats and flying foxes' (7/0.001 pbw), BIRDZILLA (referred to an eagle) (7/0.001 pbw), FISHZILLA (referred to an aggressive fish species) (7/0.001 pbw), MUMZILLA/MOMZILLA (7/0.001 pbw), RATZILLA (5/0.001 pbw), BRIDESMAIDZILLA (4/0.001 pbw), DOGZILLA (4/0.001 pbw), FOODZILLA 'nickname of a chef obsessed with her kitchen's cleanliness' (4/0.001 pbw), CATZILLA (3/0.0007 pbw), HOGZILLA (3/0.0007 pbw), SHARKZILLA 'dangerous shark' (3/0.0007 pbw), BRANDZILLA (2/0.0005 pbw), BRIDEZILLA-GROOMZILLA (2/0.0005 pbw).

Corpus data shows that -*zilla* is a morpheme added to nouns (especially, types of animal or family members) to indicate 'an overbearing person or an aggressive species'. Although its frequency is more limited than that of -(*a*)*holic*, this element has undergone a secretion process allowing for the shift from *Godzilla* to a more general 'imposing or violent example of its kind', whose aggression is reminiscent of a large dinosaur-like monster. A metaphorical extension is in *brandzilla* referring to a 'powerful brand on the market'. Moreover, -*zilla* words tend to reproduce the prosodic pattern of the full form *Godzilla* in the paradigmatic substitution: e.g., *hog*, *boss*, *mom*, and *dog* are all 'god' replacements in the analogical proportion.

New -*zilla* words and their meanings are predictable because of the analogy with the above series (*Hogzilla*, *Bosszilla*, *Bird-zilla*, *Bridezilla*, *groomzilla*, *mom-zilla*), which originally functioned as a concrete model for new words. Diachronic study and corpus investigation, however, demonstrate that -*zilla* is going in the direction of abstraction and productivity.

4.3 *Case studies 3 and 4:* DOCU- *and* -UMENTARY

The word *documentary* provides two "blend's parts" – i.e. *docu-*, as in *docudrama* [1961], and -*umentary*, as in *mockumentary* [1965] – which are not given as main entries in the OED. This dictionary describes the two blends as respectively coming from *docu-*(mentary) + *drama* 'a documentary drama' and *mock* + (doc)-*umentary* 'a film, television programme, etc., which adopts the form of a serious documentary in order to satirize its subject'. Therefore, it does not recognise the status of the two splinters as new morphemes, although it also records comparable cases created by analogy with *docudrama*:

- *docutainment* [1978] 'a film which seeks both to inform and to entertain', *docusoap* [1979]/*docu-opera* [n.d.] 'a television documentary series following people in a particular location over a period of time' (OED); see also *documusical* [1974], *docuhistory* [1981], and *docurecreation* [1983] in Green (1991: 77);

or with *mockumentary*:

- *rockumentary* [1969] 'a documentary on the subject of rock music', *shockumentary* [1970] 'a documentary film with shocking subject matter', *soapumentary* 'a

documentary series dealing with domestic situations and frequently characterized by melodrama and sentimentality' [1993] (OED).

A corpus linguistic analysis confirms the frequency and use of both splinters. COCA offers the following examples of blends beginning in *docu-* or ending in *-umentary*:

- DOCUDRAMA/DOCU-DRAMA (185/0.35 pmw), DOCUSERIES/DOCU-SERIES (23/0.04 pmw), DOCU-OPERA (4/0.007 pmw), DOCUCOMEDY (3/0.005 pmw), DOCU-SOAP (3/0.005 pmw), DOCU-MUSICAL (2/0.003 pmw).

- MOCKUMENTARY (38/0.07 pmw), ROCKUMENTARY (10/0.01 pmw), COPUMENTARY 'documentary on police action' (1/0.001 pmw), DOGUMENTARY (1/0.001 pmw), SCHLOCKUMENTARY 'documentary on inferior material' (1/0.001 pmw).

   Results from NOW for the same splinters are comparable:

- DOCUDRAMA/DOCU-DRAMA (1,508/0.39 pbw), DOCU-SERIES/DOCUSERIES (863/0.22 pbw), DOCU-SOAP/DOCUSOAP (94/0.02 pbw), DOCU-FICTION/DOCUFICTION (56/0.01 pbw), DOCU-REALITY (50/0.01 pbw), DOCU-FILM/DOCUFILM (45/0.01 pbw), DOCU-COMEDY (17/0.004 pbw), DOCU-MOVIE (9/0.002 pbw).

- MOCKUMENTARY/MOCUMENTARY/MOCK-UMENTARY (995/0.26 pbw), ROCKUMENTARY/ROCK-UMENTARY/ROCUMENTARY (119/0.031 pbw), DOGGUMENTARY/DOGUMENTARY (18/0.004 pbw), BRICKUMENTARY (11/0.002 pbw), SHOCKUMENTARY (11/0.002 pbw), VLOGUMENTARY (9/0.002 pbw), BLOCUMENTARY/BLOCKUMENTARY 'documentary on Lego building blocks' (6/0.001 pbw), DONUTUMENTARY (3/0.0007 pbw), SCHLOCKUMENTARY (3/0.0007 pbw), SPOCKUMENTARY (3/0.0007 pbw), etc.

It is worth noting that both splinters do not involve any secretion process, only abbreviation. In other words, they provide the meaning 'documentary' to the blends including them: while *docu-* coordinates with other genres, such as film/movie, drama, comedy, etc., the first element in *-umentary* words determines the type of documentary (e.g. on rock music, on shocking matters, on dogs, etc.). From the phonological viewpoint, *-umentary* words share rhyming or quasi-rhyming initial parts (*mock*, *block*, *schlock*, *spock*; *dog*, *vlog*), which increase the similarity between the new words and their models (*rock-* and *shock-umentary*) and encourage the formation of still novel words according to the same pattern.

4.4 *Case study 5:* -EXIT

The fifth case study (*-exit*) is neither attested in the OED nor in COCA as a formative one. The independent word is of course attested, but its use in blends has for long been disregarded by lexicographers, and only the NOW corpus offers pertinent examples, with the country's name functioning as abbreviated initial element given in brackets:

- FREXIT (from France, 135/0.03 pbw), NEXIT (from the Netherlands, 88/0.02 pbw), AUSEXIT/AUS-EXIT (from Austria, 27/0.007 pbw), DEXIT (from DE for 'Germany'/from Denmark, 25/0.006 pbw), ITALEXIT (from Italy, 22/0.005 pbw), IREXIT (from Ireland, 21/0.005 pbw), SWEXIT (from Sweden, 20/0.005 pbw), SPEXIT (from Spain, 16/0.004 pbw), EIREXIT (from Eire, 11/0.002 pbw), AMEXIT (from America, 9/0.002 pbw).

The OED has recently included *Brexit* and *Grexit* as neologisms formed "by compounding" *Gr*(eek) and *Br*(itish) with *exit*, in spite of the evident blending process intervening here. Although the quotes of both are from 2012 onwards, OED3 specifies that *Grexit* was originally the model for *Brexit*, now far more common than *Grexit*. Similarly, in *Wordspy*, these are the only examples recorded.

Results from corpus analysis show that *-exit* has not acquired productivity yet, but it might become an independent morpheme in the future, due to the specification process it undergoes when added to a country's name (i.e. 'withdrawal from the EU').

5. Discussion of results

The analysis of some cases of splinters, combining forms, and secreted affixes has helped us in their understanding and classification. According to the analysis conducted here, these are very close phenomena, and the same element may even represent different phenomena at different stages. For instance, in an earlier version of the OED, *-(a)holic* was labelled "combining form", while in the current version it is considered a "suffix", which has even been used as a converted noun (*olic*, *aholic*). By contrast, *-exit* has been included in the OED as a blend's part only since March 2017, when the updated version has recorded *Grexit* and *Brexit* as entries. Before that date, it was only included as an independent word.

While an actual and precise labelling of these elements is possible only synchronically, diachronically we can remark that for most of them there is an evolution, with various intermediate stages determining their changes from non-morphemic segments to actual morphemes. Originally, these elements were parts of blends, or SPLINTERS, which often merged with other word parts. For instance, the original splinter status of *-holic* is demonstrated by the fact that it was initially combined with other splinters (e.g. *choco*(late) or *carbo*(hydrates)), and *-zilla* similarly merged with a word part in *thespzilla*.

Splinters, or blend's parts, also have the characteristic of not involving reinterpretation, but mere abbreviation. Thus, for instance, *docu-* or *-umentary* contribute to the novel words *docuseries* or *dogumentary* the same meaning as the full word *documentary*. Thus, on the one hand, they are repeatedly used to obtain new words, both neologisms (*docusoap*, *rockumentary*) and occasionalisms (*docu-reality*, *donutumentary*), but, on the other, they have not acquired the abstraction of secreted forms.

SECRETED COMBINING FORMS, or SECRETED AFFIXES, by contrast, involve a secretion process, which often entails a semantic generalisation or, more infrequently, a specification process. For instance, *-(a)holic* is no longer connected to 'alcohol', but generally denotes 'a person addicted to what is specified by the first element'. In the same way, *-zilla* is a secreted form entailing semantic reinterpretation and referring to 'a particularly imposing person or thing'. That is, *-zilla* has lost its connection with the character of Godzilla and only retained some of its semantic features (e.g. violence, strength, aggression, or dangerousness).

14

Specification has instead occurred when the element *-exit* has been used in words such as *Grexit* and *Brexit*, specifically referring to 'the exit from the European Union or Eurozone'.

Abbreviation or secretion, however, are not the only discriminating factors helping distinguish splinters from secreted combining forms or secreted affixes. Frequency and productivity are additional criteria. That is, only when a splinter becomes frequently used and allows for abstraction, it can be considered an established combining form. Combining forms can be also mere abbreviations of longer words with no new meaning, such as *Br-* from *British*, or *Euro-* from *European*, but it is their regularity in use to determine the real productivity of these forms. *Euro-*, for instance, is a recognised ABBREVIATED COMBINING FORM, whereas *Br-* has acquired some regularity only in recent times, but only in the news and in a very limited number of words.

In fact, a corpus linguistics analysis of all the elements examined in section 4 has shown that they are not moderate in productivity, but highly productive and frequent (e.g. *shopaholic*, *dogzilla*, *docucomedy*, *dogumentary*, COCA; *shareaholic*, *groomzilla*, *docufiction*, *brickumentary*, *Frexit*, NOW). Needless to say, different degrees of productivity (vs. creativity) are displayed by secreted affixes/combining forms vs. abbreviated combining forms vs. splinters, the latter being less regular than the others, and hence representing the earliest step in the development of a rule.

6. The productivity of splinters, combining forms, and secreted affixes

Overall, the productivity of abbreviated or secreted forms such as splinters, combining forms or secreted affixes depends on three main factors:

1) the availability of a series sharing the same formation,
2) its potential to become a schema model for the creation of new words, and
3) the extent to which this schema is actually exploited in language use (i.e. profitability).

As far as availability is concerned, these forms are made available thanks to a process of "reinterpretation" (Hock 1991: 176), "reanalysis" (Hopper & Traugott 2003: 56), or morphological re-segmentation. For instance, *God-zilla*, *Water-gate*, and *ex-ercise* have been reanalysed as complex forms, whereas *docu-mentary*, *doc-umentary* (vs. *document* + *-ary*), *alco-holic* (vs. *alcohol* + *-ic*), and *eco-logical* (vs. *ecolog*(y) + *-ical*) have undergone morphological re-segmentation. Other comparable series involving reanalysis or re-segmentation include:

- *-kini* ← *bikini* [1947] (from the name of the atoll *Bikini*), reanalysed as *bi-* + *kini*, as in *monokini* [1964] 'a one-piece swimming costume' (OED3), *trikini* [1967] 'ladies' swimsuit which consists of three main areas of fabric' (OED2) (cf. *tankkini* [n.d.], Lehrer 2007: 131; *burkini* [2014], *facekini* [2017], *WWW*).
- *-wich* ← *sandwich* (from John Montagu, 4th Earl of *Sandwich*), reanalysed as *sand* + *wich*, as in *duckwich* [1943] and *turkeywich* [1943] (Algeo 1991: 6), *bagelwich*, *croissanwich* [n.d.] (Lehrer 2007: 123).
- *-furter* ← *Frankfurt-er* (from *Frankfurt*), re-segmented as *frank* + *furter* (Marchand 1969: 213), as in *krautfurter* [1949] 'a frankfurter topped or stuffed with sauerkraut' (OED3), *chickenfurter*, *shrimpfurter* [n.d.] (Lehrer 2007: 120).

Re-segmentation often occurs also in wordplays or puns. For instance, on *Facebook*, users have shared a post where the word *geometry* (originally from the two neoclassical combining forms *geo-* and *-metry*) was reinterpreted as having a final verb *try* (as *geome-try*, with a different pronunciation /αI/), humorously replaced by near-homophonous words in the paradigmatic substitution: i.e. *geome-cry*, *geome-why*, *geome-bye*, and *geome-die*. In this case, the splinter *geome-* was not meant to become available for new formations, but merely created for a textual (humorous) function, to obtain funny occasionalisms, not intended neologisms.

As for the relevance of a schema model in the formatives analysed in this study, from the morphological viewpoint, both frequent splinters and combining forms (or secreted affixes) belong to paradigmatic morphology, in that they trigger a schema model, based on similarity with concrete prototype words. Analogy via schema, however, excludes some of the examples included by Warren (1990) among "combining forms" (e.g. *-tro* in *outro* [1967] 'a concluding section' OED3, after *intro*(duction)) or by Bauer et al. (2013) among "splinters" (e.g. *-o* in *speako* [2001] 'an error in speaking', after *typo*(graphical error)). In the model of analogy adopted here (see § 2.1), *outro* is a case of surface analogy, with a unique model *intro*, whereas *-o* has limited productivity (see *writo* [1993], *thumbo* [2009] in *Wordspy*, but unattested in the OED or in corpora such as COCA/NOW).

As far as profitability is concerned, some splinters (e.g. final *-burger*, *-ercise*, *-gate*, *-gram*, *-holic*, *-kini*, *-speak*, *-ware*; initial *docu-*, *eco-*) have become regular, productive, and are frequently reused for still novel formations. Thus, they illustrate cases where profitability works at its best. Others remain surface analogies (e.g. *-tro*), or are limited in productivity (e.g. *-o* 'an error' vs. slang *-o*; cf. *-wich*, *-furter*, which are not as productive as *-burger*), and still others (*geome-*) are simply used once, in wordplays, and die as soon as they have been created.

Therefore, the former productive splinters have acquired a morpheme status, become secreted affixes or (secreted/abbreviated) combining forms and triggered analogy via schema. The latter have remained unproductive splinters used in surface analogy.

Some such cases still continue to be borderline. For instance, there is not much consensus throughout linguistic and lexicographic works as to whether elements such as *-alicious* (from *delicious*, in *babelicious* [1991] 'sexually attractive' OED3), *-rific* (from *terrific*, as in 'What a *brillerific* comp!' [1989] OED3), *-tainment* (from *entertainment*, in *infotainment* [1980] 'informative material presented in an entertaining way' OED3), *-tastic* (from *fantastic*, in *poptastic* [1992] 'excellent' OED3), etc. should be considered "combining forms" (OED) or "splinters" (Lehrer 2007, Bauer et al. 2013). According to the analysis conducted here, their generalisation and frequency in use suggest their inclusion among fully transparent combining forms.

Another case generating divergent opinions or contrasting analyses is *-arian*. In line with the OED, *nutarian* [1909] 'a vegetarian whose diet is based on nut products', *breatharian* [1979] 'a person who consumes no nutrients other than those absorbed from the air', and *fruitarian* [1893] 'one who lives on fruit' are obtained from the "suffix" *-arian*, like *vegetarian*. However, in the same dictionary, *flexitarian* [1998] 'a person who follows a primarily but not strictly vegetarian diet' (OED3) is analysed as a blend from *flexi*(ble + vege)*tarian*. This latter analysis denies the morpheme status of *-(t)arian*, which Bauer et al. (2013: 526–527) instead consider to be a "splinter" (their cover term for both blend's parts and secreted combining forms/affixes) generally referred to 'someone with a diet restriction'. This controversy could be explained through diachronic and semantic motivations. That is, while *flexitarian* contains a splinter *-(t)arian* and can be analysed as a blend, *nutarian*, *breatharian*, and *fruitarian* contain

a secreted suffix referring to 'someone with a diet restriction', after reanalysis of *vegetarian* as *vege*(table) + -(*t*)*arian*, and a secretion process which attributes a more general meaning to the latter suffix. A blend analysis is indeed impossible in many -(*t*)*arian* words. For instance, a *breatharian* is not 'a vegetarian', and in *meatarian* cited by Lehrer (2007: 126) the meaning of *meat* is even semantically incompatible with the sense of *vege*(table) in *vegetarian* [1842] 'a person who abstains from eating animal food' (OED3). Thus, in these examples, -(*t*)*arian* corresponds to the concept of secreted combining form or secreted affix, where the process involved is secretion rather than mere abbreviation.

7. Conclusion

This study has shown that the shift from a splinter to a combining form, or even to a secreted affix, is a diachronic issue. As Lehrer (2007: 121) observes: "[s]ince there is a scale from a completely novel splinter to a completely conventional morpheme, the transition from splinter to independent morphemehood is a diachronic process".

In this diachronic process, independent morpheme status is generally accompanied by a semantic process of generalisation (e.g. -(*a*)*holic* 'a person who appears to be addicted to the thing, activity, etc., expressed by the first element' OED3) or specification (e.g. -*exit* 'withdrawal from the EU'), which allows for the abstraction of secreted forms. When there is no abstraction, the splinter remains an abbreviated combining form (-*tainment*, -*umentary*; *docu*-), or a word part in blends.

Hence, the analogical model adopted here is gradual and envisages different stages for (and different types of) analogical formation. Diachronically, there were key phases in which the profitability of an element increased, the element became a fully productive, transparent morpheme and originated a series that acted as schema. Those crucial nodes determined the shift from surface analogy, with a unique model (e.g. *sugarholic* after *workaholic*), to analogy via schema, with a series of actual words as model (e.g. *workaholic*, *cake-aholic*, *chocoholic*, *foodaholic*, etc. working as model for *computerholic* or *sexaholic*). Therefore, analogy via schema may be viewed as the first step towards the development of a rule, from concrete words as model to an abstract rule-format template.

Secretion and the ensuing generalisation/specification also allow for the formation of many occasionalisms which, in spite of their ephemerality, contribute to stabilise the pattern of -(*a*)*holic* or similar formations, in their journey from extra-grammatical (blend) to marginal morphology (combining form), or even to standard grammatical rule (suffix).

As Klégr & Čermák (2010: 235) claim, "[a]nalogy is the backbone of creativity, i.e. the native speaker's ability to extend the language system in a motivated but unpredictable (non-rule governed) way which may or may not subsequently become rule-governed, predictable and productive". Frequent splinters, combining forms, and secreted affixes show this evolution from motivated but unpredictable to productive and (partially) predictable. Analogy helps creativity in this process towards productivity.

Appendix A. Secreted splinters, combining forms, and secreted affixes mentioned in the paper

Secreted	Origin	Examples
-(a)holic 'a person addicted to'	alcoholic	chocoholic, tobaccoholic, workaholic
-alicious 'embodying the qualities denoted by the first element to a delightful or attractive degree'	delicious	babelicious, bootylicious, groovalicious
-(t)arian 'someone with a (diet) restriction'	vegetarian	breatharian, fruitarian, nutarian
-burger 'patty served on a bun'	Hamburger	beefburger, cheeseburger, chickenburger
-exit 'exit from the European Union'	exit	Brexit, Grexit, Spexit
-furter 'a (highly seasoned) sausage'	Frankfurter	chickenfurter, krautfurter, shrimpfurter
-gate 'an actual or alleged scandal'	Watergate	Billygate, Dallasgate, Monicagate
-gram 'a message delivered by a representative of a commercial greetings company'	telegram	Gorillagram, kissogram, Rambogram
-kini 'beach garment worn by women'	Bikini	burkini, monokini, trikini
-rific 'embodying the qualities denoted by the first element in an excellent way or to an extreme degree'	terrific	brillerific, yoga-rific (body), splatterific (film)
-speak 'a characteristic mode of speaking'	Newspeak	computer-speak, royalspeak, techspeak
-tastic 'designating someone or something perceived as excellent or remarkable'	fantastic	choketastic, gaytastic, poptastic
-wich 'an article of food for a light meal with a savoury filling'	Sandwich	bagelwich, duckwich, turkeywich
-zilla 'a particularly imposing person or thing'	Godzilla	Bosszilla, Bridezilla, mom-zilla

Appendix B. Abbreviated splinters, combining forms, and secreted affixes mentioned in the paper

Abbreviated (initial/final)	Origin	Examples
Br-	Britain/British	Bremain, Brentry
docu-	documentary	docudrama, docusoap, docu-opera
e-	electronic	e-journal, e-publication, e-text
eco-	ecological	eco-damage, eco-problem, eco-tragedy
Euro-	European	Eurofashion, Euro-Japanese, Euromarket
nega-	negative	negademand, negatrip, negawatt
-ercise	exercise	boxercise, dancercise, sexercise
-tainment	entertainment	docutainment, edutainment, infotainment
-umentary	documentary	mockumentary, rockumentary, shockumentary
-ware	software	courseware, freeware, shareware

References

Adams, Valerie. 1973. *An introduction to Modern English word-formation*. London: Longman.

Algeo, John (ed.). 1991. *Fifty years among the new words: A dictionary of neologisms, 1941-1991*. Cambridge: Cambridge University Press.

Arndt-Lappe, Sabine & Plag, Ingo. 2013. The role of prosodic structure in the formation of English blends. *English Language and Linguistics* 17(3). 537–563.

Baldi, Philip & Dawar, Chantal. 2000. Creative processes. In Booij, Geert E., Lehmann, Christian, Mugdan, Joachim, Kesselheim, Wolfgang & Skopeteas, Stavros (eds.), *Morphologie-Morphology: An international handbook of inflection and word-formation*. Vol. 1. 963–972. Berlin/New York: Mouton de Gruyter.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie, Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Benczes, Réka. 2006. *Creative compounding in English*. Amsterdam/Philadelphia: Benjamins.

Christofidou, Anastasia. 1994. *Okkasionalismen in poetischen Texten*. Tübingen: Narr.

COCA: *Corpus of Contemporary American English*. 1990–2015 (http://corpus.byu.edu/coca/) (Accessed 2017-01-05).

Crawford Camiciottoli, Belinda B. 2015. 'All those Elvis-meets-golf-player looks': A corpus-assisted analysis of creative compounds in fashion blogging. *Discourse, Context and Media* (http://dx.doi.org/10.1016/j.dcm.2015.10.002) (Accessed 2015-01-06).

Doleschal, Ursula & Thornton, Anna M. (eds.). 2000. *Extragrammatical and marginal morphology*. München: Lincom Europa.

Dressler, Wolfgang U. 1993. Word-formation: Poetic licence. In Ronald E. Asher & Simpson, James M. Y. (eds.), *The encyclopedia of language and linguistics*. Vol. 9. 5028–5029. Oxford: Pergamon Press.

Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. In Ursula Doleschal & Thornton, Anna M. (eds.), 1–10.

Dressler, Wolfgang U., Mayerthaler, Willi, Panagl, Oswald & Wurzel, Wolfgang U. 1987. *Leitmotifs in natural morphology*. Amsterdam/Philadelphia: Benjamins.

Facebook (https://www.facebook.com/) (Accessed 2017-01-05).

Fradin, Bernard. 2000. Combining forms, blends and related phenomena. In Ursula Doleschal & Thornton, Anna M. (eds.), 11–59.

Franceschi, Daniele. 2013. The *cum*-construction in present-day English. *Italian Journal of Linguistics / Rivista di Linguistica* 25(2). 221–260 (http://linguistica.sns.it/RdL/25.2/Franceschi.pdf) (Accessed 2017-01-05).

Green, Jonathon. 1991. *Neologisms. New words since 1960*. London: Bloomsbury Publishing.

Gries, Stefan Th. 2004. Isn't that fantabulous? How similarity motivates intentional morphological blends in English. In Michel Achard & Kemmer, Suzanne (eds.), *Language, culture, and mind*, 415–428. Stanford: CSLI Publications.

Gries, Stefan Th. 2012. Quantitative corpus data on blend formation: Psycho- and cognitive-linguistic perspectives. In Vincent Renner, Maniez, François & Arnaud, Pierre (eds.), *Cross-disciplinary perspectives on lexical blending*, 145–168. Berlin/New York: Mouton de Gruyter.

The Guardian Online. 2017. Guardian News and Media Ltd (http://www.theguardian.com/uk) (Accessed 2017-01-05).

Hock, Hans H. 1991. *Principles of historical linguistics*, 2nd ed. Berlin/New York: Mouton de Gruyter.

Hohenhaus, Peter. 1996. *Ad-hoc-Wortbildung. Terminologie, Typologie und Theorie kreativer Wortbildung im Englischen*. Frankfurt am Main: Peter Lang.

Hopper, Paul J. & Traugott, Elizabeth Closs. 2003. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

The Independent Online. 2017 (www.independent.co.uk/) (Accessed 2017-01-05).

Kemmer, Suzanne. 2003. Schemas and lexical blends. In Hubert C. Cuyckens, Berg, Thomas, Dirven, René & Panther, Klaus-Uwe (eds.), *Motivation in language: From case grammar to cognitive linguistics. Studies in honour of Günter Radden*, 69–97. Amsterdam/Philadelphia: Benjamins.

Kilani-Schoch, Marianne. 1988. *Introduction à la morphologie naturelle*. Bern: Peter Lang.

Kilani-Schoch, Marianne & Dressler, Wolfgang U. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.

Klégr, Aleš & Čermák, Jan. 2010. Neologisms of the 'on-the-pattern-of' type: Analogy as a word-formation process? In Martin Procházka, Malá, Markéta & Šaldová, Pavlína (eds.), *The Prague school and theories of structure*, 229–241. Göttingen: V&R unipress.

Köpcke, Klaus-Michael. 1993. *Schemata bei der Pluralbildung im Deutschen: Versuch einer kognitiven Morphologie*. Tübingen: Gunter Narr Verlag.

Köpcke, Klaus-Michael. 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language* 25. 293–319.

Ladányi, Mária. 2000. Productivity, creativity and analogy in word formation (WF): Derivational innovations in Hungarian poetic language. In Gábor Alberti & Kenesei, Istvàn (eds.), *Approaches to Hungarian: Papers from the Pécs conference*, vol. 7, 73–90. Szeged: JATEPress (http://ladanyi.web.elte.hu/derivational_innovations.pdf) (Accessed 2013-01-08).

Lehrer, Adrienne. 1996. Identifying and interpreting blends: An experimental approach. *Cognitive Linguistics* 7(4). 359–390.

Lehrer, Adrienne. 2003. Understanding trendy neologisms. *Italian Journal of Linguistics / Rivista di Linguistica* 15(2). 369–382.

Lehrer, Adrienne. 2007. Blendalicious. In Judith Munat (ed.), *Lexical creativity, texts and contexts*, 115–133. Amsterdam/Philadelphia: Benjamins.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*, 2nd ed. München: Beck.

Marle, van, Jaap. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Mattiello, Elisa. 2007. Combining forms and blends: The case of *scape*. In Laura Jottini, Del Lungo, Gabriella & Douthwaite, John (eds.), *City-scapes: Islands of the self. Language studies*. Vol. 2. 115–130. Cagliari: CUEC.

Mattiello, Elisa. 2013. *Extra-grammatical morphology in English. Abbreviations, blends, reduplicatives, and related phenomena*. Berlin/Boston: Mouton de Gruyter.

Mattiello, Elisa. 2016. Analogical neologisms in English. *Italian Journal of Linguistics / Rivista di Linguistica* 28(2). 103–142.

Mattiello, Elisa. 2017. *Analogy in word-formation. A study of English neologisms and occasionalisms*. Berlin/Boston: Mouton de Gruyter.

Mattiello, Elisa & Dressler, Wolfgang U. Forthcoming. The morphosemantic transparency/opacity of novel English analogical compounds and compound families.

Milizia, Denise. 2014. In, out, or half way? The European attitude in the speeches of British leaders. *Lingue e Linguaggi* 11. 157–175.

Miller, Gary D. 2014. *English lexicogenesis*. Oxford: Oxford University Press.

Motsch, Wolfgang. 1981. Der kreative Aspekt in der Wortbildung. In Leonhard Lipka (ed.), *Wortbildung*, 94–118. Darmstadt: Wissenschaftliche Buchgesellschaft.

NOW: *News on the Web Corpus*. 2010–today (http://corpus.byu.edu/now/) (Accessed 2017-01-05).
OED2: *Oxford English Dictionary Online*, 2nd ed. Oxford: Oxford University Press (http://www.oed.com/) (Accessed 2017-01-05).

OED3: *Oxford English Dictionary Online*, 3rd ed. Oxford: Oxford University Press (http://www.oed.com/) (Accessed 2017-01-05).

Plag, Ingo. 1999. *Morphological productivity. Structural constraints in English derivation*. Berlin/New York: Mouton de Gruyter.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Ronneberger-Sibold, Elke. 2000. Creative competence at work: The creation of partial motivation in German trade names. In Ursula Doleschal & Thornton, Anna M. (eds.), 85–105.

Ronneberger-Sibold, Elke. 2008. Word creation: Definition – function – typology. In Franz Rainer, Dressler, Wolfgang U., Kastovsky, Dieter & Luschützky, Hans C. (eds.), *Variation and change in morphology. Selected papers from the 13th international morphology meeting*, Vienna, February 2008, 201–216. Amsterdam/Philadelphia: Benjamins.

Štekauer, Pavol. 2000. *English word formation: A history of research (1969–1995)*. Tübingen: Gunter Narr Verlag.

Štekauer, Pavol & Lieber, Rochelle (eds.). 2005. *Handbook of word-formation*. Dordrecht: Springer.

Storrs, Carina. 2017. You may be a pizza-holic: Research says some foods addicting. CNN (http://edition.cnn.com) (Accessed 2017-01-05).

Urban Dictionary. 1999–2017 (http://www.urbandictionary.com/) (Accessed 2017-01-05).

Warren, Beatrice. 1990. The importance of combining forms. In Wolfgang U. Dressler, Luschützky, Hans C., Pfeiffer, Oskar E. & Rennison, John R. (eds.), *Contemporary morphology*, 111–132. Belin/New York: Mouton de Gruyter.

Wordspy – The Word Lover's Guide to New Words. 1995–2017 (http://www.wordspy.com/) (Accessed 2017-01-05).

Zwicky, Arnold M. & Pullum, Geoffrey K. 1987. Plain morphology and expressive morphology. In Jon Aske, Beery, Natasha, Michaelis, Laura & Filip, Hana (eds.), *Proceedings of the thirteenth annual meeting of the Berkeley Linguistics Society, general session and parasession on grammar and cognition*, 330–340. Berkeley: Berkeley Linguistics Society.

Dr. Elisa Mattiello, Ph.D.
Department of Philology, Literature and Linguistics
University of Pisa
Via S. Maria 67 – 56126 Pisa (Italy)
E-mail elisa.mattiello@unipi.it
Phone +39 050 2215863

Anglicisms and German near-synonyms.
What lexical co-occurrence reveals about their meanings

Heike Baeskow, Bergische Universität Wuppertal & Jürgen Rolshoven, Universität zu Köln

In this article, subtle semantic and connotative contrasts between the well-established anglicisms Dealer, Drink, Song, Skyline and shoppen and German near-synonyms will be identified for the first time on the basis of their distribution in large quantities of text. Co-occurrence matrices automatically generated from three journalistic corpora and two web-based corpora will show that despite a certain degree of semantic overlap, the anglicisms convey specific meaning components or connotative nuances which make them suitable for semantic and/or connotative differentiation. Semantic information retrieved from co-occurrence profiles will be mapped onto qualia structures, which are part of Pustejovsky's (1996) Generative Lexicon and help to model the generic knowledge associated with the referents of lexical items. Connotations, which are not part of the semantic representation, are supposed to comprise stylistic, emotional and communicative-pragmatic information as well as the Nebensinn ('by-sense') of a lexical item.

Keywords: *anglicisms, contrastive analyses, lexical co-occurrence, corpus linguistics, qualia structures, connotations*

1. Introduction

It is a well-known fact that borrowing is driven not only by naming need or the need to fill lexical gaps in the receptor language. Quite frequently, a loan word displays a meaning component or a connotative nuance that is not inherent to a semantically similar native word and thus allows for lexical differentiation. Anglicisms in German are a case in point. For example, while lexical items like *Quad*, *beamen* or *smart* were required to denote concepts for which German lacked adequate expressions, loan words such as *Drink*, *Song* or *shoppen* semantically correspond with *Getränk*, *Lied*, and *einkaufen*. Nevertheless, these pairs are not completely synonymous because there are very subtle contrasts.[1]

Although there is a vast amount of literature on anglicisms in German, contrastive analyses are restricted to the studies of Yang (1990), Kettemann (2006) and Onysko & Winter-Froemel (2011) so far, and apart from Yang's attempt to describe semantic contrasts in terms of word-field theory, a theoretical framework is missing (cf. Baeskow (2018) for an overview of these studies). Moreover, computer-based corpus analyses are still at a very early stage in this field. While English loan words and the contexts in which they occur had to be manually collected and saved on filing cards by the time Yang wrote his thesis on anglicisms in the German newsmagazine *Der Spiegel*, Kettemann's analyses benefit from COSMAS II, a linguistic corpus provided by the IDS Mannheim, whose concordances allow him to describe the distribution of *cool*, *shoppen*, *Event* and their native near-equivalents *kühl*, *einkaufen* and *Ereignis* in German newspapers. The analyses performed by Onysko & Winter-Froemel (2011),

---

[1] As pointed out for example by Weisgerber (1962: 167), complete synonymy is generally avoided.

which aim at a pragmatic classification of anglicisms in German, are based on *Spiegel 2000* – an electronic resource comprising *Spiegel* editions from 1994 up to 2000.[2] While these studies – like the present article – focus on anglicisms and German near-synonyms, Sosnizka (2014) compares collocational patterns of the source and target language in electronically accessible German and American business and news magazines by using *WordSmith Tools*. Her analyses follow the tradition of British contextualism (e.g. Firth 1957, Jones & Sinclair 1974).

The aim of this study is to present results of innovative contrastive analyses performed on the basis of lexical co-occurrence in large quantities of text. The database for this study, which will be described in detail in section 2, comprises three journalistic corpora (the Spiegel-, Focus- and Zeit-Archiv) and two web-based corpora (Twitter and the German version of WaCky). Proceeding from co-occurrence matrices generated from these corpora for selected anglicisms and potential German equivalents, it will be argued that semantically relevant lexical items which are among the 30 most frequent co-occurrences[3] of a key word in at least two corpora used for this study reflect some generic knowledge we associate with the concept the key word denotes and thus help to identify lexical contrasts as well as semantic overlap. Furthermore, it will be shown that despite a certain degree of semantic overlap, the anglicisms convey specific meaning components or shades of meaning and thus allow for semantic and/or connotative differentiation. The search for semantically relevant information in the co-occurrence matrices was guided by the notion of *qualia structure* (henceforth abbreviated as QS), which is an essential component of Pustejovsky's (1996) Generative Lexicon.

The article is structured as follows: Section 2 provides an overview of the Generative Lexicon and the connotative framework. In section 3, the corpora as well as the methods developed for the contrastive analyses will be introduced, and it will be shown that qualia structures are a suitable device for information retrieval. In section 4, representative contrastive analyses of *Dealer* vs. *Händler*, *Drink* vs. *Getränk*, *Song* vs. *Lied*, *Chanson*, *Arie*, *Skyline* vs. *Stadtsilhouette* and *shoppen* vs. *einkaufen* will be presented. The article ends with a conclusion in section 6.

2. The theoretical framework

Referring to the Aristotelian philosophy, Moravcsik (1975) emphasizes that the ideas we have about objects, processes etc. are based on understanding, which goes beyond mathematical propositions and generic knowledge. Understanding a concept involves understanding that it is part of the extra-linguistic reality, where it interacts with other concepts. Moravcsik's (1981) account of understanding is intensional and suggests that

---

[2] This reference corpus is also used by Onysko (2007) for a comprehensive analysis of anglicisms in German.

[3] This notion will be applied here to any of the 30 lexical items that occur most frequently in the context of a given key word, i.e. of an anglicism or its German near-synonym. Given this rather broad conception of the linguistic context, the much and controversially discussed term 'collocation' (cf. Sosnizka 2014, chap. 2 for a historical overview) will be restricted to fixed collocations, i.e. to unpredictable combinations of words which have to be learned and memorized as units and which affect the levels of both syntax and word-formation, e.g. *Er raucht stark* "He smokes heavily", *ein starker Raucher* "a heavy smoker" (Lipka 2002: 182f).

the capacity to identify extensions is not part of the speakers' linguistic competence. For example, although speakers generally know the concept DISEASE, they are not necessarily able to identify instances of it without consulting a physician. While the extension remains constant, the intensional representation of concepts is incomplete, or, more precisely, it allows for variation. In particular, it is enriched from childhood to adolescence, and further change may be due to new scientific insights. Just like an adult's intensional representations of concepts differ from those of a child, the knowledge of an expert is more specific than that of the layperson. Variation as to intensional knowledge is possible even among different generations of scientists. Nevertheless, conceptual representations that differ in specificity are not mutually exclusive, but complementary, and incompleteness does not necessarily constitute an obstacle to communication.

2.1 *From Aristotelian aitiae to qualia structures*

In the Aristotelian philosophy, understanding a concept means knowing its *aitiae*, i.e. the four generative factors which define its distinctive properties, constituency, function, and coming into existence. These factors provide a common scheme for the representation of intensional knowledge – independently of its specificity. The essence of the aitiational framework is that semantic competence is measured not by the ability to distinguish an entity x from all other entities in the world, but by the degree to which an individual is able to distinguish the entity x from some family of elements *y*, *z*, *w* – i.e. from elements of the same semantic field. Thus, semantic competence involves for example the ability to distinguish bees from wasps, hornets etc. or to differentiate between verbs of motion such as *run*, *walk*, or *gallop* (Moravcsik 1981: 22–23). This assumption makes the aitiational approach a solid basis for contrastive analyses.

In Pustejovsky's *Generative Lexicon* (henceforth abbreviated as GL), the ensemble of generative factors, which are referred to as 'modes of explanation' (1996:76), makes up the qualia structure of a concept.[4]

- FORMAL: that which distinguishes it in a larger domain;
- CONSTITUTIVE: the relation between an object and its constituent parts;
- TELIC: its purpose and function;
- AGENTIVE: factors involved in its origin or "bringing it about".

Qualia structures, which provide a multi-dimensional representation of concepts in the form of well-defined types and relational structures (Pustejovsky 1996: 78), interact with three other levels of representation: an argument structure, an (extended) event structure, and a lexical inheritance structure. As far as argument structures are concerned, Pustejovsky (1996: 63–64) distinguishes four types of arguments, namely TRUE ARGUMENTS, which are syntactically realized, DEFAULT ARGUMENTS, i.e. "parameters which participate in the logical expressions in the qualia, but which are not

---

[4]   It should be noted that qualia structures do not necessarily specify all of these generative factors. For example, since natural types such as stone or water are not inherently associated with a function, the TELIC quale remains unspecified (unless some purpose or intention is being imposed on these concepts); cf. Pustejovsky (2003: 375–379).

necessarily expressed syntactically", e.g. *John built a house out of bricks* (1996: 63), SHADOW ARGUMENTS, which are semantically incorporated into the lexical item (e.g. *to butter*), and TRUE ADJUNCTS, which allow for temporal or spatial modification and which are not part of a lexical item's semantic representation.

As pointed out already by Moravcsik, qualia structures, which define the prototypical meaning components of concepts, constitute a suitable device for representing fine-grained contrasts between semantically similar lexical items. For example, according to Pustejovsky (1996: 77–79), the English nouns *novel* and *dictionary* are semantically related (i.e. both objects are books in a general sense). Nevertheless, they differ not only with respect to their content, but also with respect to their purpose (we *read* a novel, but we *consult* a dictionary) and their coming into being (i.e. a novel is *written*, whereas a dictionary is *compiled*). These differences, which are reflected by collocational patterns, are represented at the CONSTITUTIVE, TELIC and AGENTIVE quale respectively, as illustrated below.

(1)

$$
\textbf{\textit{novel}}
$$

ARGSTR =
ARG1	= y: physical_object
D-ARG1	= x: human
D-ARG2	= z: human

QUALIA =
FORMAL	= y
CONSTITUTIVE	= narrative (y)
TELIC	= read (e, x, y)
AGENTIVE	= write (e, z, y)

(2)

$$
\textbf{\textit{dictionary}}
$$

ARGSTR =
ARG1	= y: physical_object
D-ARG1	= x: human
D-ARG2	= z: human

QUALIA =
FORMAL	= y
CONSTITUTIVE	= information (y)
TELIC	= consult (e, x, y)
AGENTIVE	= compile (e, z, y)

Both *novel* and *dictionary* have an argument (y) in their argument structure that is of the ontological type 'physical_object' and corresponds to the referential argument <R> introduced by Williams (1981) for nouns. The argument structures additionally display

two default arguments D-ARG1 and D-ARG2 of the type 'human', whose variables (x) and (z), like (y), are bound in the qualia structures. In (1), (x) is related to a reading event which defines the function of the concept NOVEL, whereas (z) actively participates in the writing event that causes the coming into existence of (y). Thus, (x) and (z) correspond to the reader and the author respectively. The structure of the text associated with novel is specified in the CONSTITUTIVE quale. The qualia value 'narrative' contrasts with the value 'information', which defines the internal structure of a dictionary. Further contrasts emerge from the TELIC and AGENTIVE quales, where (x) is the agent of a consulting event and (z) actively participates in an event of compiling (y).

Qualia structures have been designed to account for context-specific interpretations of lexical items. For example, a sentence like *Mary began the novel* is perfectly interpretable in GL theory because the contextual sense 'began reading' is inferable from the reading event in the TELIC quale of *novel* – the complement of the polysemous verb *to begin* (cf. (1)). This mechanism is referred to as true complement coercion by Pustejovsky (1996: 115–117).

As pointed out above, the dynamics of intensional knowledge does not generally affect the exchange of information. According to Moravcsik (1981: 24), "there is enough overlap among the aitiational schemes of speakers to make communication, in most cases, possible." If speakers share information as to aitiational schemes (or qualia structures in Pustejovsky's terminology), we should expect that indicators of this common intensional knowledge are retrievable from corpora.

Evidence for this assumption comes from computational linguistics and empirical ontology. An early attempt to extract qualia-related information from an electronic corpus is made by Anick & Pustejovsky (1990). Their search is based on "collocational patterns" which linguistically reflect the content of qualia structures. For example, the collocations *read a book*, *read a tape*, *use the mouse* or *with the mouse* are suggestive of the TELIC role associated with the concepts BOOK, TAPE, or MOUSE. Anick & Pustejovsky's results were designed for information retrieval applications.

More recent works have shown that an automatic extraction of qualia relations from large corpora is possible as well. In particular, Cimiano & Wenderoth (2005), whose intention is to reveal the impact of qualia structures on natural language processes and to support the lexicographer's work, developed a system which automatically produces the qualia structures (it acquired by learning) from the World Wide Web. Poesio & Almuhareb (2008) attempt to evaluate the relevance of attributes extracted from the Web for the description of concepts, with attributes being conceived of as essential properties of concepts as defined in AI, linguistics and philosophy – including qualia roles.

Although the aims of these works are quite different from the aims of the present article, it becomes clear that aitiational schemes anticipated by Moravscik (1981: 17) "as a psychological and semantic claim" are an integral part of language use rather than theoretical constructs. The case studies in section 4 it will show that qualia-related information is also traceable in co-occurrence matrices and helps to formalize semantic contrasts between anglicisms and German near-synonyms.

2.2 *Non-denotative information*

Lexical items do not only have a referential function, but also convey rather heterogeneous facets of meaning which are traditionally referred to as connotations. Following Erdmann (1900), Ludwig (1986, 2002), Yang (1990: 45–46) and Fries (2007) it is assumed here that connotations comprise (a) all kinds of associations which make up the 'Nebensinn' ('by-sense') of a word, (b) its 'Gefühlswert' ('emotional value'), (c) stylistic information, and (d) communicative-pragmatic information.

If associations generally arise in the context of a lexical item, i.e. if they are not formed in the minds of individual speakers, they are part of its connotative information content and thus constitute its 'Nebensinn' in the sense of Erdmann (1900: 82). Erdmann illustrates this by-sense for the pair *Krieger* "warrior" and *Soldat* 'soldier'. While the former is associated with fight and battle, the latter rather evokes the image of barracks and parade ground. Additionally, the noun *Krieger* is archaic and only of historical relevance.

Emotions are defined by Fries (2007: 298) as clusters of subjective-psychological experience and motor behaviour which are encoded in linguistic or indexical signs. As far as the emotional predisposition (abbreviated as EM) is concerned, Fries (2007: 309) introduces three independent dimensions, namely (a) emotional polarity $EMpol+$, $EMpol-$, $EMpol_0$, (b) emotional expectation $EMexp-$, $EMexp+$, and (c) emotional intensity $EMint+$, $EMint-$. Although this classification was developed for German linguistic signs such as *leider* 'unfortunately', *hoffentlich* 'hopefully', *pfui* 'ugh', *Bewunderung* 'admiration' etc., it is also suited to describe the emotional potential of anglicisms. Emotional values are either word-inherent (e.g. *Wellness* $_{EMpol+}$, *stalken* $_{EMpol-}$ "to stalk", *Wow!*$_{EMexp-}$) or unfold in the context. For example, the anglicism *Crash* has a pejorative connotation in colloquial speech, where it refers to a collision, or in the jargon of Stock Exchange, but it is an emotionally neutral technical term in linguistics. Lexical items are predisposed to occur in particular domains of communication, and their use may be determined by the social norm. Lexicographically, Ludwig (1986: 187–193) distinguishes between the communicative predisposition, which determines stylistic properties at the word-level (e.g., *Gesicht* 'face' is stylistically neutral, whereas *Antlitz* 'countenance' and *Visage* 'mug' range above and below 'neutral' respectively), and communicative-pragmatic markers, which assign lexical items to specific domains of communication (e.g. 'technical', 'group-specific'), signal a speaker's or author's emotional attitude, or predict temporal or regional restrictions for the use of a lexeme (e.g. 'archaic', 'Afro-American English').

In German, anglicisms are often used for stylistic or communicative-pragmatic reasons. Stylistic analyses of anglicisms in German have been strongly influenced by Galinsky (1963). An important notion is that of colouring ('Kolorit'), which is comparable to timbre in music. If the author of a text uses an anglicism in order to evoke a particular setting, to characterize a certain group of speakers or depict a new trend, local, social, or technical colouring[5] is being created (e.g. *Sheriff*, *cool*, *Wellness-Shop*). These stylistic devices enable him or her to convey a more vivid impression of his/her subject (cf. Pfitzner 1978, chapter III, Yang 1990, chapter 4, Donalies 1992:

---

[5]   The corresponding German notions are 'Lokalkolorit', 'Sozialkolorit', and 'Fachkolorit'.

104–106, Leutloff 2003, Götzeler 2008: 282–289). In German youth language, anglicisms are primarily associated with social functions, which are discussed in detail by Androutsopoulos (1998, chapter 7).

3. Computerlinguistic methods and their application

As far as research on anglicisms is concerned, the studies described in this article are innovative in two ways. First, the contrastive analyses are performed within a well-established theoretical framework, and secondly, the relevant data were extracted from very large heterogeneous corpora. The texts were accessed via a searchable database (KANG) comprising ten subcorpora, which was built by Jürgen Rolshoven and his colleagues from the University of Cologne. Within the scope of a research project on anglicisms in German, co-occurrence matrices for anglicisms and semantically similar German words were automatically generated from five subcorpora – the Spiegel-, Focus- and Zeit-Archiv, Twitter, and WaCky. The Spiegel-Archiv (documented period: 1946–2015, data volume: 1.7 GB) is a collection of texts from the German newsmagazine *Der Spiegel*, which has always been a source of lexical innovation (cf. Carstensen 1965: 22, Onysko 2007: 98). A more recent newsmagazine is *Focus*, which was first published in 1993. The Focus-Archiv (data volume: 375 MB) comprises articles from the first edition up to 2015. The Zeit-Archiv (documented period: 1946–2016, data volume: 1.3 GB) is based on articles from the supra-regional German newspaper *Die Zeit*. The data for these three subcorpora were extracted using the open online access to the archive of each of the papers. Apart from the articles' raw text, the publishing year and month as well as the name of the source were saved in the corpus as meta data. The Twitter corpus (2011–2015), whose tweets were collected via the Twitter-API[6], was kindly provided by Prof. Dr. Chris Biemann (University of Darmstadt, now University of Hamburg) and Dr. Martin Riedl (University of Darmstadt). It is regularly updated and used for further processing at the University of Cologne. After filtering it for relevant German texts it now consists of approximately ten million tweets with a data seize of 3.5 GB including some minor meta data such as the publishing date. KANG also provides access to the German version of the Web-As-Corpus Kool Yinitiative (WaCky), which was designed by Marco Baroni and colleagues[7] to crawl the web. The German WaCky corpus represents the largest part of the data (9.5 GB) used to build KANG.[8] All the data were standardized and merged into the searchable database.

The co-occurrence matrices were generated separately for each corpus by the computer linguists from the University of Cologne. By generating co-occurrence matrices from journalistic texts and web-based corpora it was ensured that the co-occurrences are not text-type specific. Each matrix displays a hierarchy of thirty lexical items that occur most frequently in the context of (a) an anglicism and (b) a potential German equivalent in a given corpus. The following extract of the co-occurrence matrix

---

[6] https://dev.twitter.com/streaming/reference/get/statuses/sample
[7] http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf
[8] A collection of working papers on the Web as Corpus was published by Bernadini & Baroni (2006). It is online available at http://wackybook.sslmit.unibo.it/.

for the key word *Drink* from the Spiegel-Archiv is intended to provide a first impression.

Table 1: The first 15 co-occurrences of *Drink*

```
Kookkurenzen von DRINK  (Spiegel-Archiv)
=====================================
||||||||||||||||||    BAR   Häufigkeit: 34   Anteil: 0.06938775510204082
|||||||||||    SOFT   Häufigkeit: 19   Anteil: 0.03877551020408163
|||||||||||    ERSTEN        Häufigkeit: 18   Anteil: 0.036734693877551024
||||||||||    HAND  Häufigkeit: 14   Anteil: 0.02857142857142857
|||||||||    COLA   Häufigkeit: 12   Anteil: 0.024489795918367346
|||||||||    UHR   Häufigkeit: 12   Anteil: 0.024489795918367346
|||||||||    NEHMEN       Häufigkeit: 12   Anteil: 0.024489795918367346
|||||||||    ENERGY       Häufigkeit: 12   Anteil: 0.024489795918367346
||||||||    MANN  Häufigkeit: 11   Anteil: 0.022448979591836733
||||||||    ALKOHOL       Häufigkeit: 10   Anteil: 0.02040816326530612
||||||||    WHISKY       Häufigkeit: 10   Anteil: 0.02040816326530612
||||||||    ESSEN Häufigkeit: 10   Anteil: 0.02040816326530612
||||||||    LETZTEN       Häufigkeit: 10   Anteil: 0.02040816326530612
||||||||    ZEIT   Häufigkeit: 10   Anteil: 0.02040816326530612
|||||||    PROZENT       Häufigkeit: 9   Anteil: 0.018367346938775512
             […]
```

The co-occurrence values were determined by counting the occurrences of all the words in a certain context window around the term and sorting them by frequency. Semantically irrelevant words (stop words, i.e. functional items, as opposed to lexical items) and compounds were removed from the texts before applying the context window. The window breadth was fixed to ±5, i.e. the span to be considered consists of five words to the left and five words to the right of a key word. The given values are the ratios of found contexts containing the respective co-occurrence.

The ranking of the co-occurrences in terms of co-occurrence values is less relevant for the present study because it is assumed here that looking for common co-occurrences in the five corpora more efficiently contributes to the identification of concept-defining information. The basic criterion is that a word appears among the "Top 30" co-occurrences as described above, independently of its position in the matrix. As observed by Baroni & Lenci (2008), token frequency may be accidental or result from fixed expressions. Their analyses have shown for example that the fixed expression *year of the tiger* occurs much more frequently in their corpora than the pattern *tail* and *tiger*, which signals a semantic relation and allows for different realizations (e.g. *tail of the tiger*, *tigers have tails*, *tigers with tails* etc.). Thus, the analyses performed in this study are qualitative rather than quantitative.

Co-occurrence matrices deliver the raw material for knowledge representations. They do not immediately reveal recurrent patterns such as qualia relations or other pieces of lexical information, but each matrix contains clues as to the interpretation of its key word. If these clues recur in at least two co-occurrence matrices from the

corpora used for this study, we may assume that they are lexically relevant. The search for useful lexical information in the co-occurrence matrices was primarily guided by the qualia-based approach.[9] Co-occurrences, which also include word-forms (e.g. *singt*, *sang*, *gesungen* in the case of *Lied* 'Song'), may be suggestive of qualia relations, qualia values, ontological types, hyponyms, arguments, fixed collocations, idioms, or connotative information.

Lexical items that contribute to the definition of a key word were manually extracted from the co-occurrence matrices and make up the co-occurrence profile[10] of the key word. While a co-occurrence matrix also contains irrelevant information, a co-occurrence profile is a subset of all the co-occurrences automatically identified for a key word which only contains interpretable 'slices' of knowledge.

Co-occurrences may be irrelevant for different reasons. For example, unlike *Bar*, *Soft* and *Energy* are irrelevant for the semantics of *Drink* (cf. Table 1) because they are part of the highly frequent compounds *Soft Drink* and *Energy Drink*. Although compounds were automatically excluded from the co-occurrence analyses, these sequences could not be prevented from entering the matrices because they do not follow the German orthographic convention to write compounds in one word (i.e. *Softdrink*, *Energydrink*). Despite their relative frequency, co-occurrences such as *ersten* 'first', *Uhr* 'clock' or *Mann* 'man' are irrelevant as well because they cannot be systematically related to the concept DRINK.

Co-occurrence profiles provide an appropriate starting point for the construction of knowledge representations and for the performance of contrastive analyses because they reveal contrasts as well as semantic overlap between loan words and semantically similar native equivalents. Significantly, competing near-synonyms do not imply lexical redundancy, but signal a need for semantic or connotative differentiation, which may be very subtle.

4. Contrastive analyses

Loan words frequently undergo meaning specification when they enter the receptor language. The reason is that borrowing takes place at the level of *parole*. In a concrete communicative situation, a particular meaning component of a loan word is required rather than its entire meaning spectrum. Thus, speakers of the receptor language tend to borrow a lexical unit, i.e. a form and one meaning component related to this form (cf. Yang 1990: 46, 167, Gévaudan 2002: 25, Onysko 2007: 16–17, Winter-Froemel 2011: 213). However, as observed by Carstensen (1965: 256), English loan words – especially older ones – do not preserve a static meaning in German, but rather undergo semantic change. In particular, they are subject to meaning extension, which manifests itself most obviously in compounding. In Baeskow (2018), the well-integrated loan words *Drink*, *Dealer* and *Job* were analysed as to their behaviour in the head position of hybrid N+N compounds, which were automatically extracted from the Spiegel- and

---

[9]   This approach also supplements generic knowledge that is not retrievable from co-occurrence matrices.

[10] This notion was borrowed from Belica (2011), who uses it in a slightly different way. According to Belica, a co-occurrence profile of an object comprises all the quantitative results of a co-occurrence analysis performed for this object.

Zeit-Archiv and manually selected from Cosmas II, the linguistic corpus provided by the IDS Mannheim[11]. It was shown that the specific meaning the head-forming nouns were associated with when they were borrowed from English may be contextually overridden by German modifiers and that meaning extension by modification coincides with a semantic approximation of the head-forming anglicisms and German near-synonyms, e.g. *Händler* 'trader', *Getränk* 'drink' and *Beruf* 'profession' in the case of *Dealer*, *Drink* and *Job*. A natural next step was to analyse the semantic behaviour of selected anglicisms beyond compounding. The basic question was whether there is a general tendency for anglicisms to extend their meaning in the course of time, so that the contrasts which distinguished them from their German equivalents are gradually blurred. Apart from two of the previously analysed nouns, namely *Dealer* and *Drink*, the analyses include the verbal anglicism *shoppen* examined by Kettemann (2006) in rather limited contexts, and two nominal anglicisms which have not yet been subject to contrastive analyses, namely *Song* and *Skyline*.

4.1 *Händler vs. Dealer*

According to the *Anglizismen-Wörterbuch* compiled by Carstensen & Busse (1993–1996)[12], the agent noun *Dealer*, which is polysemous in English, was borrowed with the very specific meaning component "someone who sells illegal drugs".

In Baeskow (2018) it was shown that this meaning component, which is also prevalent in compounds of the type *Drogendealer*, *Rauschgiftdealer* (both meaning 'drug dealer') or *Heroindealer*, may be overridden by native modifiers. Examples are *Autodealer* "car dealer", *Knoblauchdealer* 'garlic dealer', *Klingeltondealer* 'ring-tone dealer' or *Plattendealer* 'record dealer', which are not (or only metaphorically) related to the drug scene. Although compounds like these are stylistically marked so far, they suggest that the meaning of *Dealer* at least partially overlaps with the meaning of a native agent noun that has a more general meaning, namely *Händler* 'trader', but does this local, i.e. context-specific approximation also hold beyond compounding? Is it justified to state that *Dealer* has undergone semantic extension in the receptor language? Let us begin by looking at the co-occurrence profile generated for *Händler*, which provides significant clues as to the generic knowledge associated with this concept.

Table 2: Co-occurrence profile of *Händler*

HÄNDLER	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
corpus freq.	9381	3208	3671	1009	10.148
Kunden	345	269	139	31	386
Kunde		50			151
verkaufen	287	86	100	16	210

verkauft	163	65	75		183
Markt	214	69	88	11	153
Ware	186	65	79		183
Preis	171	80	55		142
Preise	175	67	57		108
Geschäft	203	62	56		
Geschäfte	113		50		
Geld	158	65	60		166
kaufen	114	60	58	11	197
Käufer		81	57		210

In particular, this profile reveals an activity prototypically associated with HÄNDLER. This activity is realized by the verb *verkaufen* 'sell' and the past participle *verkauft*, which are recurrent in the corpora. Since the activity of selling defines a professional occupation, it determines the content of the TELIC quale of HÄNDLER. Moreover, since this activity requires a human agent, the referent of this person-denoting noun must be of the type 'Person (x)', which is specified in the FORMAL quale.

As far as the object of transaction is concerned, four corpora provide the collective noun *Ware* 'merchandise', which fits the position of the internal argument (y) opened by *verkaufen*. Compounds like *Obsthändler* 'fruiterer', *Autohändler* 'car dealer', *Buchhändler* 'bookseller', *Antiquitätenhändler* 'antique dealer', *Haushaltswarenhändler* 'hardware dealer' and many others show that the very general qualia value 'Ware' allows for specification. However, since the syntactic realization of (y) is not obligatory in the context of *Händler* (in Twitter it is not among the "Top 30" co-occurrences), it is considered here to be a default argument as defined in section 2. Further co-occurrences which are logically related to the concept HÄNDLER are *Kunde(n)* 'customer(s)', *Markt* 'market', *Preis(e)* 'price(s)', *Geld* 'money', and *Geschäft(e)*. The noun *Geschäft* either refers to a shop or to a transaction business. *Markt* is polysemous, too, because it either denotes a concrete or an abstract location. In the first reading, it functions as an adjunct which spatially locates the event of selling goods associated with HÄNDLER. In its abstract reading, it refers to the interplay of supply and demand in which the trader is involved. The following sentences from WaCky illustrate certain constellations which the co-occurrences can enter with their key word (italics and translation by HB).

(3) a. *Auf dem Markt* in der Fußgängerzone *verkaufen* die *Händler* Krimskrams für wenig *Geld*.
   'In the market place in the pedestrian zone, the traders are selling odds and ends for little money.'

b. Zunächst muss ein *Händler* seine *Kunden* identifizieren.

'First, a trader needs to identify his/her customers.'

  c. Zu beiden Seiten boten *Händler* ihre *Ware* lautstark an.
'Traders loudly offered their goods on both sides.'

  d. Habt ihr gehandelt oder hat der *Händler* sofort den *Preis* genannt?
'Did you bargain or did the trader quote the price immediately?'

(4) a. Wenn an den Wochenenden *auf dem Markt* rund 30 *Händler* ihre *Waren* anbieten, [...]
'When approximately 30 traders offer their goods in the market place at weekends, […]'

  b. Zusätzlich entsteht ein neuer *Markt* für *Händler* von Emissionsberechtigungen.
'Additionally, a new market for traders of emission rights is developing.'

(5) a. Viele *Händler* hatten ihre *Geschäfte* geöffnet.
'Many traders had opened their shops.'

  b. Bei uns trägt der *Händler* die Beweislast, mit wem er ein *Geschäft* eingegangen ist.
'At our place, the trader bears the burden of proof with whom he has entered into a business.'

These findings suggest that automatically generated co-occurrence matrices do not display arbitrary information. The qualia-based framework facilitates the evaluation of co-occurrences and helps to establish semantic relations between these items and a given key word. In particular, it is a useful guide to the identification of contrasts between semantically similar key words. The following profile indicates that six co-occurrences, namely *Geld* 'money', *Geschäft* 'business', *Kunden* 'customers' *Ware* 'merchandise', *verkaufen* 'sell' and *verkauft* 'sold' are shared by the concepts HÄNDLER and DEALER. Thus, the referents of both nouns are readily identifiable as participants in a transaction event.

Table 3: Co-occurrence profile of *Dealer*

DEALER	*Spiegel*	*Focus*	*Zeit*	*Twitter*	*WaCky*
corpus freq.	1.608	454	320	316	2.386
Drogen	97	20	13	23	186
Droge			10	4	33
Polizei	71	42	17	10	115
Heroin	62	19	13		77
Stoff	65	16	8	6	44
Geld	55	9	9		41
Kokain	43	23		4	34

Junkies	40	12			29
Süchtige	30	12			
Szene	37	11			
Geschäft	36	10			
Rauschgift	35	9			27
Kunden	34	8			27
Fahnder	32	10			
Gramm	34	11			28
kaufen				6	26
Ware	28				29
verkauft				5	36
verkaufen					37

Even more importantly, Table 3 provides a reliable answer to the question whether *Dealer* underwent meaning extension beyond compounding. Apart from instances of semantic overlap, the co-occurrences suggest that *Dealer* has preserved the specific meaning component it was associated with at the time it was borrowed. Although *Ware* "merchandise" occurs in both profiles, the object of transaction is definitely restricted to drugs in the case of *Dealer*. While *Ware* is displayed only in the Spiegel-Archiv and WaCky, *Drogen* is robustly represented in each of the five corpora. In addition, there are various instantiations of this qualia value, namely *Heroin*, *Stoff* 'dope', *Kokain*, and *Rauschgift*. Even the customers to whom the drugs are sold are further specified by the nouns *Junkies* and *Süchtige* 'addicts'.

Moreover, following Pustejovsky's (1996: 229–230) distinction between 'role defining' nominals (e.g. *physicist*, *linguist*, *violinist*) and 'situationally-defined' nominals (e.g. *pedestrian*, *passenger*, *student*), it is argued here that *Händler* is interpreted generically and that *Dealer* has a specific interpretation. While *Händler* constitutes an occupational title, a person referred to as *Dealer* is identified as such only if he is engaged in selling illegal drugs. The terms 'role-defining' nominals and 'situationally-defined' nominals reflect the more established distinction between *individual-level nominals* (ILNs) and *stage-level nominals* (SLNs)[13]. While nouns of the former type define the role of an individual independently of the activity performed at the time of reference, the interpretation of nouns of the latter type require the actual performance of characteristic activities. In GL theory, this difference is accounted for by assigning the activities typically associated with ILNs and SLNs to the TELIC and the AGENTIVE quale respectively, as exemplified below for the concepts HÄNDLER and DEALER.

---

[13]   The distinction between stages and individuals goes back to Carlson (1977)

(6) $\text{QS}_{\text{HÄNDLER}}$ = $\begin{bmatrix} \text{FORMAL} = & \text{x: Person} \\ \text{TELISCH} = & \text{verkaufen (e, x, y: Ware, z: Kunden)} \end{bmatrix}$

(7) $\text{QS}_{\text{DEALER}}$ = $\begin{bmatrix} \text{FORMAL} = & \text{x: Person} \\ \text{AGENT.} = & \text{verkaufen (e, x, y: Drogen, z: Junkies)} \end{bmatrix}$

Note that the co-occurrence profile displayed in Table 3 also shapes the connotative information content of *Dealer*. The co-occurrence of this key word with nouns referring to drugs, with *Polizei* 'police', *Fahnder* 'investigator', *Szene* 'scene', and with the nouns *Junkies* and *Süchtige* (both of which are inherently specified for <EMpol–> in the sense of Fries 2007) suggests that the activity specified in the AGENTIVE quale is illegal. Here we are dealing with a connotation or *Nebensinn* "by-sense" in Erdmann's (1900: 82) terminology, which is not inherent to *Händler* and which evokes negative emotions.[14]

The results presented in this section have shown that there is a partial overlap between the concepts associated with HÄNDLER and DEALER. Both nouns refer to transactions involving merchandise, customers and money, but the meaning of the latter is more specific because the prototypical object of transaction is specified in its qualia structure and conveys the impression of illegality (along with other co-occurrences which are not part of the profile of HÄNDLER). The contrastive analyses of *Dealer* and *Händler* in compounding (Baeskow 2018) led to similar results, but also revealed that *Dealer* is quite frequently substituted for *Händler* (and related native nouns) in the head position of N+N compounds for stylistic purposes in journalistic texts. If the qualia value 'Drogen' is overridden by a modifier that does not refer to drugs, but to objects of value, harmful substances or limited resources, the connotation 'illegal' is preserved by the compound (e.g. *Grundstücksdealer* 'estate dealer', *Sprengstoffdealer* 'dealer in explosives', *Elfenbeindealer* 'ivory dealer'). If the modifier refers to everyday items or food, the compound is either metaphorically/humorously related to the drug scene, or the aspect of illegality is contextually suppressed (e.g. *Strumpfdealer* 'hosier', *Plattendealer* 'record dealer', *Sonnenbrillendealer* 'dealer in sunglasses'). Beyond compounding, however, *Dealer* and *Händler* are not (yet) exchangeable, because *Dealer* is too strongly associated with the distribution of illegal drugs. The frequent occurrence of *Dealer* in the context of *Drogen* in all five corpora suggests that this anglicism largely preserved the specific meaning it displayed when it was borrowed from English.

---

[14] Although the referent of the emotionally neutral noun *Händler* may also deal in drugs (cf. *Drogenhändler*, *Rauschgifthändler*), the range of goods to be distributed is in principle infinite.

4.2 *Getränk vs. Drink*

Like *Dealer*, the noun *Drink* was subject to meaning specification when it entered the German language in the 19th century. According to Stiven (1936: 71), it was exclusively used to refer to alcoholic beverages. In the AWB, it is primarily defined as "alkoholisches (Misch-)Getränk" ('(mixed) alcoholic beverage'). Furthermore, it is pointed out that this anglicism extended its meaning to denote not only alcoholic beverages, but also "(Misch-)Getränke jeglicher Art" ('all kinds of (mixed) beverages'). The contexts provided in this reference work suggest that instances of meaning extension in compounding (e.g. *Cola-Drinks*, *Suppen-Drink* 'soup drink', both 1978) and beyond are not attested before the late 1970s. In Baeskow (2018), new trends as to the use of *Drink* are identified. First, *Drink* has become a vogue word in the context of modifiers related to wellness and health, where it unfolds a pseudo-medicinal flavour that is not conveyed by its German near-synonyms (e.g. *Darmgesundheitsdrink* 'intestinal-health drink', *Fett-weg-Drink* 'fat-away drink', *Herzdrink* 'heart drink'). Secondly, *Drink* is preferred over *Getränk* (and similar near-synonyms) if the beverage is prepared by mixing the (alcoholic or non-alcoholic) ingredients denoted by the modifier. The mixing event, which is represented as an option in the AWB, is becoming more salient in view of the numerous innovative compounds whose modifiers refer to natural ingredients (e.g. *Erdbeer-Bananen-Drink* "strawberry-banana drink", *Kiwi-Apfel-Drink* 'kiwi-apple drink', *Karotten-Brokkoli-Drink* 'carrot-broccoli drink'). In these compounds and in contexts like those in (8), which are attested even before the 1970s, semantic approximation of *Drink* and the native noun *Getränk* is most evident.

(8) a. Sie verabreichte siebzehn Kraftfahrern *Drinks*. Acht erhielten reinen Fruchtsaft, die anderen ein alkoholisches Getränk. (Zeit-Archiv, 6/1952)
     'She gave drinks to seventeen drivers. Eight of them got pure fruit juice, the others an alcoholic drink.'

   b. So nahm er in der Milchbar des Bundeshauses einen *Drink* mit einem Parlamentsportier [...] (Spiegel-Archiv, 4/1956)
     'Thus, he had a drink with the parliamentary porter in the milk bar of the federal parliament building […]'

In (9), *Mineralwasser* 'mineral water' is obviously referred to as *Drink* for stylistic purposes. If a perfectly appropriate native noun (in this case *Getränk*) is contextually replaced by an anglicism in order to render an everyday concept more attractive or prestigious, we are dealing with an instance of semantic-stylistic upgrading in the sense of Pfitzner (1978: 194–195). In journalese, this stylistic device may be accompanied by a touch of irony.

(9) a. *Mineralwasser* ist der absolute *In-Drink* der Saison. - Dem Durstigen, der im Café des "Steigenberger Parkhotels" zu Hamburg ein Mineralwasser bestellt, reicht der Ober eine Karte: die Wasserkarte. (Focus-Archiv, 8/1994)
     'Mineral water is the absolute in-drink of the season. - The waiter passes a card to the thirsty guest who orders mineral water at the café of the 'Steigenberger Parkhotel' in Hamburg: the water menu.'

b. Schon aus Solidarität mit Daniel Schreiber kann ich zu beiden Büchern nur einen *Drink* empfehlen: *Mineralwasser*. Sage niemand, das sei fantasietötend: Es gibt Mineralwasser zu äußerst fantasievollen Preisen. (Fokus-Archiv, 2/2015)
'If only in solidarity with Daniel Schreiber I can recommend just one drink for both books: Mineral water. Don't tell anyone this is killing fantasy. There is mineral water available for most fanciful prices.'

Despite these instances of semantic approximation, the co-occurrence matrices of both nouns, from which compounds were excluded, indicate that the basic contrast is still prevalent. The co-occurrence matrix generated for *Getränk* gave rise to the following profile:

Table 4: Co-occurrence profile of *Getränk*

GETRÄNK	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
Cola	47	6	26	10	74
Coca	28		15		
Bier	29	13	27	21	140
Wasser	25	10	16	27	52
Kaffee	22	10	23	19	94
Tee	8		11	7	113
Wein	25	16	18		153
Whisky	19		14		
Champagner	9		11		
Alkohol	16	4	8		
alkoholisches		5		14	
Flasche	22		13		
Glas	14	6	8		70
trinken	12	8	16	23	105
getrunken	8				

To begin with, *Getränk* is of the ontological type 'Flüssigkeit' ('liquid'). This very general piece of information is not directly retrievable from the co-occurrence profile, but it is reflected by the co-hyponyms *Cola*, *Bier* 'beer', *Wasser* 'water', *Kaffee* 'coffee', *Tee* 'tea', *Wein* 'wine', *Alkohol*, *alkoholisches* 'alcoholic', *Whisky*, and *Champagner* 'champagne'. Further support for the typing of *Getränk* as 'Flüssigkeit' comes from the co-occurrence partners *Flasche* 'bottle' and *Glas* 'glass', which denote

containers for keeping liquids in. The function typically associated with *Getränk*, namely *trinken* 'to drink'
, is directly retrievable from the co-occurrence profile, which displays the infinitive as well as the word forms *trank* [preterite] and *getrunken*[past participle]. Significantly, the co-occurrences indicate that *Getränk* refers to all kinds of beverages, including those containing alcohol. Thus, the constituency remains unspecified in its QS.

(10)

$$
\textbf{Getränk} \\
\text{ARGSTR} = \begin{bmatrix} \text{ARG1} & = & y: \text{Flüssigkeit} \\ \text{D-ARG1} & = & x: \text{Person} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{FORMAL} & = & y \\ \text{TELISCH} & = & \text{trinken (e, x, y)} \end{bmatrix}
$$

As expected, the co-occurrence profile of *Drink* is very similar to that of *Getränk*. In particular, the referents of both nouns are of the same ontological type and share a common *telos*. Moreover, there are instances of *Whisky*, *Cola* and *Coca* in both profiles. In spite of these semantic similarities, subtle contrasts between *Getränk* and *Drink* become apparent. Consider the following co-occurrence profile:

Table 5: Co-occurrence profile of *Drink*

DRINK	*Spiegel*	*Focus*	*Zeit*	*Twitter*	*WaCky*
corp. freq	490	234	350	613	2.837
Bar	34	15	33	8	161
nehmen	12	6	26		81
nimmt	8		7		42
trinken				14	38
trinke				7	
getrunken				10	
Glas		8			64
Hand	14	7			43
Alkohol	10	5			43
Whisky	10		8		

Cola	12		9		
Coca			8		
abends	8		7		
Abend		5		8	55
Wodka	7		7		

This profile suggests that *Drink* typically refers to alcoholic beverages (beyond compounding). First of all, co-hyponyms like *Kaffee*, *Wasser* or *Tee* identified for *Getränk*, which denote common thirst quenchers, do not surface among the 30 most frequent co-occurrence partners of *Drink*. Secondly, this anglicism occurs in the context of the noun *Bar* in each of the corpora used for this study. Since it is part of the generic knowledge that a bar is a location where alcoholic beverages are served (unless the reference is restricted by a modifier, as in the compound *Milchbar* 'milk bar' (cf. (8b)), this relatively stable co-occurrence indicates that {Alkohol} continues to be a defining property which basically distinguishes *Drink* from its native competitor *Getränk*. Thirdly, the function associated with *Drink* (TELIC) is lexically realized by word forms of *nehmen* (rather than *trinken*). In German, the phrase *einen Drink nehmen* (literally 'to take a drink') is a fixed collocation in which *Drink* refers to an alcoholic beverage. Phrases like *?einen Softdrink/Energydrink/Frucht-drink nehmen* are definitely marked.

(11)

> **Drink**
>
> ARGSTR = [ARG1 = y: Flüssigkeit
> D-ARG1 = x: Person]
>
> QUALIA = [FORMAL = y
> KONST. = {Alkohol}
> TELISCH = nehmen (e,x,y)]

Further collocations in which *Drink* exclusively refers to alcohol are *auf einen Drink* 'for a drink' and *bei einem Drink* 'with a drink'. Since stop words (i.e. function words) were excluded from the automatic co-occurrence analyses, these constructions are not retrievable from Table 5, but they are listed in the AWB (entry for *Drink*), where they are associated with the image of a social gathering with friends or business partners at which alcoholic beverages are enjoyed. Similarly, a relaxing, informal atmosphere is evoked by the fixed collocation *mit einem Drink in der Hand* 'with a drink in his/her hand', for which there is an obvious indicator in the co-occurrence profile, namely the noun *Hand* (Spiegel, Focus, and WaCky). Attention should also be paid to the adverbial *abends* 'in the evening' and the noun *Abend* 'evening', which are to be found among

the co-occurrences of *Drink* in five corpora. Since these temporal relations are absent from the profile of *Getränk*, we may conclude that the social gathering typically (though not exclusively) takes place in the evening when work is done. Concrete examples are provided below:

(12) a. Wenn er *abends* ausgehen will, hat er die Wahl zwischen 25 Klubs, in denen er seinen *Drink* nehmen kann. (Spiegel-Archiv, 5/1948)
'When he wants to go out in the evening, he may chose among 25 clubs to have his drink.'

b. Er lädt sie zum Lunch und unterhält sie *des Abends* bei festlichem Dinner und *Drinks* vorm Kamin [...] (Spiegel-Archiv, 9/1987)
'He invites them for lunch, and in the evening he entertains them with a festive dinner and drinks in front of the fireplace.'

c. Ein frischer, angenehm aufmunternder *Drink* für den späten *Abend*. "Zur Bar-Kultur", sagt Daun, "gehört das richtige Getränk zur richtigen Situation." (Focus-Archiv, 7/2015)
'A fresh, pleasantly encouraging drink for the late evening. "The right drink for the right situation makes up the bar culture", says Daun.'

To sum up, the co-occurrence profile generated for *Drink* from five extensive corpora suggests that this anglicism preserved its reference to alcoholic beverages in German. However, the qualia value {Alkohol}, which distinguishes it from *Getränk*, does not have absolute character, but constitutes a prototypical semantic property which may be overridden by a modifier (e.g. *Frucht-Drink*) or the larger context (cf. (8) and (9)). In spite of the partial semantic overlap of *Drink* and *Getränk*, the anglicism provides connotative information which is not inherent to its native competitor.

4.3 *Lied vs. Song, Chanson, Arie*

Lexical co-occurrence is a suitable device not only for comparing word pairs, but also helps to identify contrasts and similarities between members of a lexical field. For example, the noun *Musik* 'music' and the Verb *singen* 'to sing' (including various word-forms such as *singt* or *sang*) are frequently represented in the context of the anglicism *Song*. These lexical items also co-occur with at least three other key words, namely with the native noun *Lied* (which corresponds with English *Song*) and two further loanwords – *Chanson* and *Arie* 'aria'. Each of these nouns is of the ontological type 'Musikstück (y)' 'piece of music' (FORMAL), and in each case, 'singen (e, x: human, y)' determines the TELIC quale. However, despite the semantic overlap, *Lied*, *Song*, *Chanson* and *Arie* are not freely exchangeable. Indicators of subtle contrasts are provided by the individual co-occurrence profiles. Let us begin by comparing *Song* with the more general native noun *Lied*.

Table 6: Co-occurrence profile of *Song*

SONG	Sp	Foc	Zt	Tw	WaC
corpus freq.	3.162	1.444	1.706	7.380	10.382
Lied	94	34	44		200
Album	70	61	61	211	418
Band	115	59	44		381
Musik	74	46	43		285
Pop	60	29			
Rock	66		26		149
schreiben	47				
geschrieben	54	37	26		158
singen	72	30	27	85	
singt	63	29	48		
sing		32		97	
Sänger	47		27		
Platte	54		35		122
heißt	110	30	55	86	161
hören			33	144	182
gehört				132	138
hört			27	79	

Table 7: Co-occurrence profile of *Lied*

LIED	Sp	Foc	Zt	Tw	WaC
corpus freq.	1.668	1.367	2.485	5.992	875
singen	122	569	320	195	62
singt	60	39	116	103	20
sang	57		54		13

sangen	36		46		
gesungen	48		75		14
Musik	26	23	46		23
Text	29		35		17
alte	35	20	90		
alten					14
heißt		22	70	156	15
Melodie	27		37		20

The first noteworthy contrast revealed by these profiles is related to the coming-into-being process, which is defined at the AGENTIVE quale. In addition to the 47 occurrences of the infinitive *schreiben* 'to write' in the Spiegel-Archiv, the past participle *geschrieben* 'written' systematically occurs in the context of *Song*. It is missing only in Twitter. Although *Lied* collocates with *schreiben*, too (*ein Lied schreiben*), this verb is not among the thirty most frequent co-occurrences of the native near-equivalent. This contrast allows for the conclusion that the concept denoted by *Song* is more readily associated with the composer (or *songwriter*, which is another anglicism in German) than the concept denoted by *Lied*. Similarly, the relatively frequent occurrence of *Band* in the context of *Song* indicates that the anglicism is more closely related to the performers than *Lied* (hence *ein Beatles-Song*, *ein Abba-Song*, *ein Song von den Stones* etc.).

Further co-occurrences which allow for a distinction between *Song* and *Lied* are *Pop* and *Rock*. These items do not contribute to the semantic structure of *Song*, but rather provide connotative information because they suggest that we are dealing here with a modern piece of music of anglo-american origin or influence, which may be of international popularity.[15] Like *Band*, *Pop* and *Rock*, the noun *Song* found its way into other languages, too. Thus, each of these anglicisms has a high international recognition value. On the other hand, *Lied* co-occurs with two word forms of the adjective *alt* 'old', namely *alten* and *alte*, in four corpora. Depending on the linguistic context, the resulting phrase either receives a literal or a metaphorical interpretation. If a piece of music for singing is old in the literal sense, it is referred to as *Lied* rather than *Song* in German, as shown below:

(13) a. Und auch sonst stimmt am *alten Lied* vom braven Landmann kaum noch eine Verszeile. (Spiegel-Archiv, 11/1977)
'And in other respects, too, hardly any verse line of the old song of the brave countryman is still correct.'

---

[15] Of course, song texts are not restricted to English. In KANG, the anglicism *Song* frequently occurs in the context of names of German bands or musicians.

    b. Und da war Amir so gerührt, daß er auf der Stelle "Ich schwebe über den Bäumen" sang, ein *altes* arabisches *Lied*. (Spiegel-Archiv, 1/1996)
      'And then Amir was so strongly moved that he instantaneously sang 'I am floating above the trees', an old Arabian song.'

A metaphorical interpretation is required for the idiom *Es ist das alte Lied* 'It's the (same) old story', which means that a well-known problem recurs frequently. In this context, *Lied* is not replaceable by *Song* (*Es ist *der alte Song*).

      The use of *Song* may also be motivated by communicative-pragmatic and stylistic considerations. In the Twitter-Korpus, it frequently occurs in the context of modifiers or intensifiers that are typical of German youth language (e.g. *krass* 'wicked, cool, awesome' in (14a), *Hammer*[16] in (14b) and *cool* in (14c)), and quite a few utterances are accompanied by emoticons, e.g.

(14) a. Noch nie erlebt, dass jemand so krass einen *Song* angeteasert hat ☺ aber hat sich gelohnt (Twitter 8/2015)
      'I never heard someone tease a song so wickedly ☺ but it was worthwhile.'

    b. der [sic!] *song* dein herz trägt felsen fand ich wahnsinnig gut. der text ist toll und der refrain ist *hammer*! (Twitter 2012)
      'I considered the song 'Dein Herz trägt Felsen' ['Your heart is bearing rocks'] to be incredibly good. The text is great and the refrain is awesome.'

    c. bitte nehmt Jedward's [HASHTAG293601174] in eure playlist auf! das ist so ein *cooler song* und verdient echt eine chance! (Twitter 11/2012)
      'Please add Jedward's to your playlist! It's such a cool song and really deserves a chance!'

Tweets like these signal that young speakers in particular consider *Song* to be less conventional or even more prestigious than *Lied*. This is what Pfitzner (1978) refers to as 'affect', which typically manifests itself in semantic-stylistic upgrading mentioned already in section 4.2.[17] Furthermore, it is noteworthy that *Song* also surfaces in a collocation in which *Lied* is less usual, namely *einen Song performen* 'to perform a song'. While *Lied* is absent from the thirty most frequent co-occurrences of *performen*, *Song* occurs in the context of this verbal anglicism in three corpora:

---

[16] The idiom *Das ist/war der Hammer* either refers to something extremely positive or extremely negative.

[17] According to Pfitzner (1978: 211), the use of anglicisms to create semantic-stylistic downgrading is relatively restricted. In Author1 (2018), this device is exemplified by expressive compounds like *Knochenjob* 'back-breaking job', *Sklavenjob* 'slave job', *Routinejob* 'routine job', *Stressjob* 'stress job', *Höllenjob* 'hell's job', *Mistjob* 'crap job', or *Billigjob* 'low-paying job' in which *Job* is usually not replaced by *Beruf* 'profession'.

Table 8: Co-occurrence profile of *performen* (extract)

PERFORMEN	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
corpus freq.	8	16	19	189	189
SONG			2	13	16
SONGS				6	9

A Google-based search restricted to German websites yields 2.010 hits for *einen Song performen* and only 279 hits for *ein Lied performen*.[18] In this collocation, *Song* unfolds a (pseudo-)technical aura which is not conveyed by *Lied* in its basic reading '(short) piece of music to be sung'. On the other hand, *Lied* constitutes a real technical term used by music experts to refer to the art form developed by Franz Schubert and continued by composers like Hugo Wolf or Richard Strauss. In this particular reading, *Lied* was borrowed into other languages such as English or Spanish.[19] This contrast illustrates the distinction between Fachsprache ('technical terminology') and Fachkolorit ('technical colouring'). While *Lied* in the sense of 'art form' denotes a musicological concept, *Song* rather constitutes a vogue word that is suited to evoke the impression of expertise.

Interestingly, *Song* seems to have largely abandoned an early meaning component. According to the AWB (entry for *Song*), this loanword is first attested in 1798, but only occurred sporadically until the beginning of the 20th century, when it was used to refer to socio-critical cabaret songs in the Brechtian style and hence assumed a very specific meaning as compared to its English model. Although the compound *Protestsong* still exists in German, there are no indicators in the co-occurrence matrices generated from the five modern corpora which point towards a critical connotation. Thus, we may conclude that the narrow reading imposed on this loanword by Brecht was abandoned in favour of a more general meaning, which entails a semantic approximation of *Lied* and *Song*. However, despite this meaning extension, *Lied* is not completely synonymous with *Song* because of its connotative shift towards a vogue word in the jargon of light music.

As indicated above, semantic overlap is observable not only for *Song* and *Lied*. Consider the co-occurrence profile of *Chanson*, which belongs to the same lexical field.

Table 9: Co-occurrence profile of *Chanson*

CHANSON	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
corpus freq.	431	62	256	22	1.352

---

18   Last access June 20th, 2017

19   For English, *the Oxford English Dictionary* (OED) also lists the hybrid compounds *lieder-singer* and *lieder-singing*. The Corpus del Español constructed by Mark Davies even provides a Spanish agent noun *liederista*.

singen	13	3	7		
singt	13		17		
Sänger	29	3			
Sängerin	35		11		40
Song		2			38
sang			8		
gesungen			7		
französische	24	4	9		40
französischen	18		7		68
Frankreich			6		37
Lied	10		13		55
Lieder					34
Musik	13	4	10		84
deutschen	11	3			35

The profile of *Chanson* shares with the profiles of *Song* and *Lied* the co-occurrences *singen* (along with related word-forms) and *Musik*, which determine the TELIC and the FORMAL quale respectively. More importantly, however, a prototypical property of CHANSON that distinguishes this concept from the concepts SONG and LIED is retrievable from the co-occurrence profile as well, namely its French origin. Considering the fact that the overall corpus frequency of *Chanson* is relatively low (in the Spiegel-Archiv, for example, there are 3162 instances of *Song*, but only 431 instances of *Chanson*), the word-forms *französische* and *französischen* of the adjective *französisch* "French" are well represented in the context of this loanword. The Zeit-Archiv and WaCky additionally display the proper noun *Frankreich* 'France'. Semantically, the French origin of the concept CHANSON is represented at the CONSTITUTIVE quale.

(15)

$$
\begin{bmatrix}
\textbf{\textit{Chanson}} \\[4pt]
\text{ARGSTR} \;=\;
\begin{bmatrix}
\text{ARG1} & = & \text{y: Musikstück} \\
\text{D-ARG1} & = & \text{x: Person} \\
\text{D-ARG2} & = & \text{z: Stimme}
\end{bmatrix} \\[20pt]
\text{QUALIA} \;=\;
\begin{bmatrix}
\text{FORMAL} & = & \text{y} \\
\text{KONST.} & = & \text{französisch (y)} \\
\text{TELISCH} & = & \texttt{singen}\ (e, x, y)
\end{bmatrix}
\end{bmatrix}
\qquad \text{“voice”}
$$
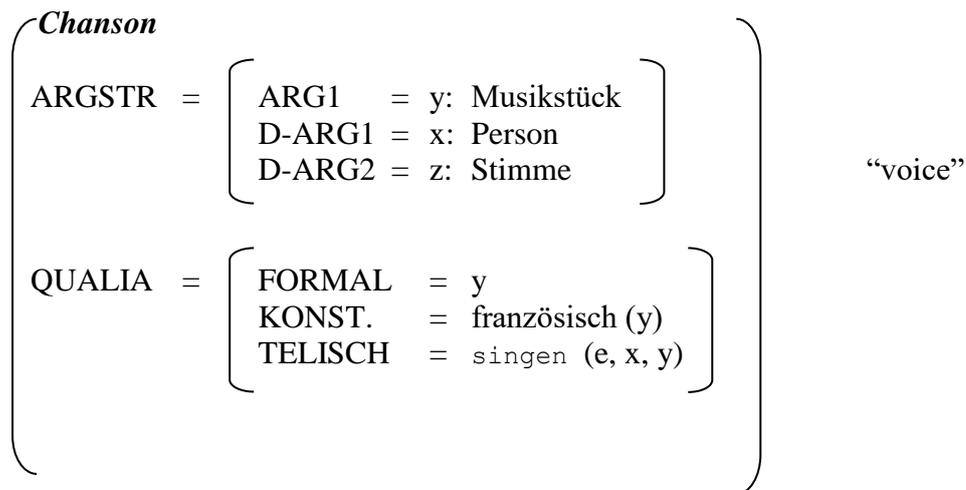
Of course, *französisch* is merely a default value, i.e. a prototypical value which may be contextually overridden. For instance, the co-occurrence profile of *Chanson* also displays the word-form *deutschen* 'German' even if this adjective is underrepresented. This finding does not come as a surprise because this type of song has also been cultivated in Germany at least since the Weimar Republic. However, although the chanson is not regionally restricted to France, it is definitely inspired by a French model (16a) and thus suitable to create local colouring (16b).

(16) a. Aznavour: Nein! Das französische *Chanson* ist ein typisch französisches Produkt, weil es die Liebe zur französischen Sprache ist, die diese Autoren [Gainsbourg, Moustaki, Piaf etc.; HB] erst hervorgebracht hat. (Zeit-Archiv 5/2014)
'Aznavour: No! The French Chanson is a typically French product because it is love for the French language which eventually brought forth these authors.'

b. Dann trällert die Empfangsdame ein *Chanson* – Erlebnisgastronomie im Stil der "Bouffes Anversois – Cafés chantants" des 19. Jahrhunderts. (Zeit-Archiv, 2/2007)
'Then the receptionist warbles a chanson – event gastronomy in the style of the 19th century "Bouffes Anversois – Cafés chantants".'

Another member of the lexical field examined in this section is *Arie* 'aria', whose distinctive property is also retrievable from the corpora.

Table 10: Co-occurrence profile of *Arie*

ARIE	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
corpus freq.	300	83	341	33	190
Oper	16	5	19		6
singen	9	6	12	3	9
singt	13	4	23	1	9
sang		4		2	
Rezitativ			15		15
Musik	5	5	5		6
Bühne	7		7		

Like *Lied*, *Song* and *Chanson*, *Arie* displays the co-occurrences *singen* (as well as related word-forms) and *Musik*, which provide the most basic components for its qualia structure. One co-occurring lexical item not shared by the other field neighbours is *Oper* 'opera'. This co-occurrence partner signals that *Arie* is a relational noun, which is typically part of an opera[20] – even if it is performed independently of the complete work. Following Bouillon et al. (2012: 1529), it is assumed here that part-of relations are represented at the CONSTITUTIVE quale, as shown in (17).

(17)

$$
\begin{bmatrix}
\textbf{\textit{Arie}} \\
\text{ARGSTR} = \begin{bmatrix} \text{ARG1} & = \text{y: Musikstück} \\ \text{D-ARG1} & = \text{x: Sänger} \\ \text{D-ARG2} & = \text{z: Stimme} \end{bmatrix} \quad \text{"voice"} \\
\\
\text{QUALIA} = \begin{bmatrix} \text{FORMAL} & = \text{y} \\ \text{KONST.} & = \text{Teil\_von Oper} \\ \text{TELISCH} & = \text{singen (e, x, y)} \end{bmatrix} \quad \text{"part\_of"}
\end{bmatrix}
$$
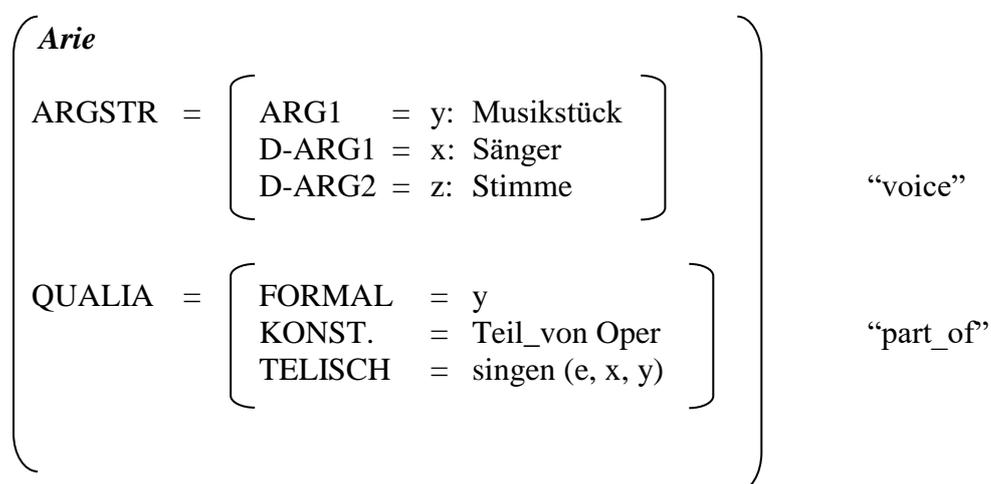
Further co-occurrences which are represented less frequently but nevertheless fit the profile of *Arie* are *Rezitativ* (Zt 15, WaC 15) and *Bühne* 'stage' (Sp 7, Zt 7). While a

---

[20] Beyond the opera, an aria is also part of a cantata or an oratorium; cf. Seeger, H. (1966): *Musiklexikon in zwei Bänden*. Leipzig: VEB Deutscher Verlag für Musik.

recitative – like an aria – is typically part of an opera (except that it is spoken instead of sung), *Bühne* is interpretable as an adjunct that spatially locates the event of singing.

While Yang's (1990: 49–55, 92–93) pre-digital identification of semes and distinctive connotative features largely depended on speakers' judgements and dictionary entries, the present analyses benefit from automatically performed co-occurrence analyses which provide clues as to the semantic and/or connotative information content of lexical items and thus help to reveal not only lexical contrasts and semantic overlap between members of word pairs, but also to determine the distribution of members of lexical fields. Moreover, since the co-occurrence matrices were generated from large quantities of text, they ensure that the relations and values that are eventually mapped onto qualia structures are empirically motivated.

4.4 *Skyline vs. Stadtsilhouette*

Although proper nouns are not part of knowledge representations, they may contribute to the connotative information content of a lexical item. As will be shown in this section, the systematic co-occurrence of a keyword with geographical names may be an indicator of local colouring ('Lokalkolorit'). Evidence comes from the profile of the anglicism *Skyline*, which is represented below:

Table 11: Co-occurrence profile of *Skyline*

SKYLINE	*Sp*	*Foc*	*Zt*	*Tw*	*WaC*
corp. freq.	299	92	205	154	1.602
New	36	11	14	9	183
York	25	6	11	6	112
Yorker	8	3		4	35
Stadt	36	12	19	4	154
Manhattan	29	9	11	3	105
Manhattans	14	9	5		35
Blick	28	13	22	10	16
Wolkenkratzer	22		7		43
Frankfurt	8	3	12	9	105
Frankfurter	16	3			78
Hochhäuser	12	6	7		28
City	10	4	7	5	45

Significantly, the noun *Skyline*, which roughly corresponds with the less frequent German-French hybrid compound *Stadtsilhouette* (literally 'city silhouette'), co-occurs with *New*, *York*, *Yorker* and *Manhattan* in all the corpora used for this study. This constellation, which is not available for *Stadtsilhouette* (*die Skyline /?Stadtsilhouette von Manhattan*)[21], locates the concept SKYLINE in a typically American setting and makes it a suitable device for local colouring – a stylistic nuance which according to Galinsky (1963: 101) conveys "the impression of American 'atmosphere' on the German listener's or reader's mind." The following sentences from the print media illustrate this effect:

(18) a. Am Fenster zieht die *Skyline* von Manhattan vorbei. Sein Ziel ist downtown, eine Bank hat ihn zum Lunchvortrag eingeladen. (Spiegel 4/2006)
   'The Skyline of Manhattan passed the window. His destination is downtown, a bank invited him for a lunchtalk.'

   b. Die *Skyline* definierte nicht nur die Stadt, sondern auch die Menschen, die zu ihren Füßen leben. Wolkenkratzer wirken plötzlich zerbrechlich. (Focus 3/2002)
   'The skyline defined not only the city, but also the humans living at its feet. Skyscrapers suddenly appear fragile.'

In (18a), the American atmosphere is conveyed by an accumulation of three anglicisms: *Skyline*, *downtown*, and the hybrid compound *Lunchvortrag*. In (18b), the noun *Skyline* has a strongly symbolic function because it is representative of life in Manhattan before and after 9/11. Both sentences would be less vivid and less authentic if *Skyline* was replaced by *Stadtsilhouette*.

Table 11 indicates that the connotative relation between *Skyline* and *New York* or *Manhattan* is transferable to megacities displaying a similar architecture. In particular, *Skyline* co-occurs with *Frankfurt*, whose financial district with all its skyscrapers is strongly reminiscent of Manhattan. On account of this similarity, Frankfurt is humorously referred to as *Mainhattan* (i.e. Manhattan on the Main). The names of other megacities only occur sporadically in the context of *Skyline*, as the following frequencies show: *Hongkong* (Sp 9, Zt 4), *Londoner* (Sp 6), *London* (Sp 6), *Dubai* (Foc 4), *Shanghai* (Tw 3*), Chicago (*WaC 23).

The link for the connotative transfer is provided by two common nouns which regularly co-occur with *Skyline* in the corpora, namely *Wolkenkratzer*[22] and *Hochhäuser* 'high-rise buildings'. Although *Hochhäuser* (unlike *Wolkenkratzer*) also co-occurs with *Stadtsilhouette* (Foc 2, Zt 3, WaCky 6), modern buildings of a considerable height do not seem to be salient defining components of this concept. Instead, *Stadtsilhouette* tends to co-occur with various nouns referring to or related to churches, namely *Frauenkirche* (Sp 1), *Kirchenschiff* 'nave' (Foc 1), *Kirchen* (Zt 1, WaC 8), and *Dom* (WaC, 8). Another co-occurrence partner is *Altstadt* "historic city" (Tw 1, WaC 6), which refers to a part of a city that usually consists of historic buildings. In the contexts provided in (19), the substitution of *Skyline* for *Stadtsilhouette* would result in stylistic markedness:

---

[21] The profile of *Stadtsilhouette* will not be displayed here because this noun has a very low corpus frequency. Instead, reference will be made to co-occurrences in the individual corpora.

[22] *Wolkenkratzer* (literally 'cloud scraper') is a loan rendition of the English noun *skyscraper*.

(19) a. Steile Staffeln, enge Gassen und spitze Giebel prägen die *Stadtsilhouette* von der Tübinger Altstadt bis hinauf zum Schloss. (Twitter 11/2012).
'Steep steps, narrow lanes and pointed gables determine the skyline of Tübingen's historic city up to the castle.'

b. Aufragend aus der Masse, geben einige wenige Gebäude der *Stadtsilhouette* ihren unverkennbaren Schnitt: rechts, im Osten, der mächtige Kubus des Alcázar, der Stadtfestung, in der Mitte der Turm der gotischen Kathedrale.
'Looming out of the mass, a few buildings shape the unmistakable skyline: to the right, in the east, the mighty cube of the Alcázar, the town fortress, in the middle the tower of the Gothic cathedral.'

The only place names co-occurring with *Stadtsilhouette* in more than one corpus are *Hamburg* (Sp 2, WaC6) and *Köln* "Cologne" (Zt 2, WaC 5) – two German cities which lack skyscrapers of a dimension typically associated with the United States. Since recently, Hamburg's highest building is the Elbphilharmonie (110 metres)[23], which is still relatively small in comparison with the Empire State Building (381 metres)[24]. In Cologne, there are some high-rise buildings, too, but since its panorama is inextricably linked to the two pointed towers of the famous cathedral – the Kölner Dom – the use of *Stadtsilhouette* constitutes an option.

From a semantic point of view, however, *Skyline* and *Stadtsilhouette* are in principle exchangeable because both nouns refer to the outline of a number of buildings seen against the sky and thus exclude the shape of natural objects such as rocks, hills or trees.[25] As shown in this section, the contrast is rather connotatively motivated, and the systematic occurrence of *New York* and *Manhattan* in the context of *Skyline* allows for the conclusion that these proper nouns trigger the use of this anglicism.

4.5 *Shoppen vs. einkaufen*

Co-occurrence matrices also allow for a comparison of lexical items other than nouns. In this section, two near-synonyms analysed by Kettemann (2006: 177–178) will be revisited, namely *shoppen* and *einkaufen* 'to do the shopping'. As indicated in the introduction, Kettemann performed his analyses on the basis of concordances from COSMAS II. Although the contexts were extracted non-automatically, they make interesting predictions. As far as *shoppen* is concerned, co-occurrences like *gerne* "with pleasure", *relaxen* 'to relax', *Spaß* 'fun', *gemütlich* 'leisurely' or *Freudinnen treffen* 'meet (female) friends' suggest that this anglicism refers to a pleasant leisure activity which is not necessarily a targeted activity. In terms of the Generative Lexicon, the TELIC quale of *shoppen* remains unspecified because there is not always a concrete need for the goods to be purchased. By contrast, the co-occurrences identified by Kettemann for *einkaufen* signal that this activity is part of the household chores, which

---

[23] http://www.hamburg.de/sehenswuerdigkeiten/4401242/hochhaeuser-in-hamburg/
[24] https://de.wikipedia.org/wiki/Empire_State_Building
[25] In English, by contrast, the denotation of *skyline* also includes "the line at which the earth or a part of the landscape appears to meet the sky" (OED), e.g. *I do love horses moving slowly against a skyline of trees* (*a*1933). Once again, the borrowing process led to meaning specification.

have to be done quickly and efficiently. Examples are *müssen* 'have to' *kochen* 'to cook', *putzen* 'to clean', *günstig* 'cheap', or *Stress*. Kettemann's findings were checked against the automatically detected contexts which gave rise to the following co-occurrence profiles:

Table 12: Co-occurrence profile of *einkaufen*

EINKAUFEN	Sp	Foc	Zt	Tw	WaC
corpus freq.	2.210	871	1.534	4.060	14.956
gehen	148	64	146	554	1.566
Geld	87	21	52	48	443
Supermarkt	52	24	39		405
Kunden	57	37	36		339
billig(er)	85	24	74		
kochen		23	46	49	361
Arbeit			32		284
arbeiten			30		262

Table 13: Co-occurrence profile of *shoppen*

SHOPPEN	Sp	Foc	Zt	Tw	WaC
corpus freq.	268	297	155	2.579	2.109
gehen	31	30	20	383	353
kaufen		13	11		49
Internet	13	16	5		51
Stadt			8	49	87
gern(e)	9	10	4	45	100
Freundin(nen)			4	51	
Freunde(n)					91
Spaß				67	84
Geld	14	10	7	59	61
Klamotten			4	37	

While both verbs enter a fixed collocation with *gehen* 'to go' (*shoppen gehen*, *einkaufen gehen*)[26] and share the co-occurrence *Geld* 'money', only *shoppen* displays the positively connoted items *gern(e)*, *Spaß* and *Freundin(nen)*, *Freund(en)* in its co-occurrence profile. In the context of *einkaufen*, there are co-occurrences such as *billig(er)* 'cheap(er)', *kochen* 'to cook', *Arbeit* 'work', or *arbeiten* 'to work'. These constellations, some of which are exemplified below, confirm Kettemann's observation that *shoppen* – unlike *einkaufen* – refers to a pleasant and relaxing activity.

(20) a. *Mit meinen Freunden* gehe ich *gern* in die Stadt *shoppen*, ins Kino oder ein Eis essen. Das macht immer riesen *spaß* [sic!] und überhaupt keine Langeweile. (WaCky)

    'I like to go shopping, to go to the cinema or to eat an ice-cream with my friends[+fem]. This is always great fun and never boring.'

    b. Morgen wahrscheinlich *mit meiner besten Freundin shoppen* und stöbern bei hugendubel. Ich freu mich jetzt schon. (Twitter, 8/2011)

    'Tomorrow probably shopping with my best friend[+fem] and rummaging at Hugendubel's. I am already looking forward to it.'

(21) a. Auf der anderen Seite könnte man sagen, daß ich Künstler bin und als solcher nicht meine ganze Zeit mit *Einkaufen* und *Kochen* verbringen will [...] (Focus, 6/1995)

    'On the other hand, one could say that I am an artist and thus do not want to spend all my time doing the shopping or cooking.'

    b. „Wenn ich nach der *Arbeit* noch *einkaufen* muß, dann Kläuschen aus der Tagesstätte abholen und mit ihm spielen – das schaffe ich nicht" (Zeit 1/1990)

    'If I have to do the shopping after work, to fetch Kläuschen from the day-care centre afterwards and to play with him – I don't manage it.'

At the lexical level, this contrast is accounted for by emotional markers in the sense of Fries (2007). While *shoppen* is inherently specified for <EMpol+>, its native competitor should be considered emotionally neutral. Although *einkaufen* may be more closely related to household chores, it is in no way a pejorative verb. Following Fries, emotional neutrality will be represented by the feature <EMpol$_0$>.

    Another revealing, though less frequent co-occurrence of *shoppen* is *Klamotten*, which is colloquial for articles of clothing. Although the goods to be purchased are not restricted to clothes, it is a matter of fact that *shoppen* is generally avoided in the context of food or domestic articles in German unless a particular stylistic effect is to be achieved. For example, the use of this anglicism in the slogan *Die ganze Woche Frische shoppen* 'buy freshness throughout the week', which covers the weekly advertising leaflet of a German discounter, is clearly marked and intended to attract the customers' attention. In this case, *shoppen* is contextually assigned the connotative feature <EMexp–>, which encodes the emotion of unexpectedness.

---

[26] A slightly archaic variant listed in the AWB is *shopping gehen*.

Contrasts are also observable as to the locations in which the events denoted by *shoppen* and *einkaufen* typically take place. In the corpora, the noun *Supermarkt* "supermarket", which is associated with a large supply of goods, self-service and special offers, only occurs in the context of *einkaufen* – the activity mainly driven by need. On the other hand, the co-occurrence profile of *shoppen* displays *Stadt* "city" and – more importantly – the noun *Internet*. Although the virtual location denoted by *Internet* is also compatible with *einkaufen* (*im Internet einkaufen*), the verb *shoppen* unfolds its positive connotation even in this context and thus implies the advantages associated with online shopping (sitting comfortably in front of the computer in order to chose a particular article independently of opening hours, using the delivery service, paying by online banking etc.).

To sum up, *shoppen* differs semantically from *einkaufen* in that it denotes an activity which is not necessarily targeted. While *einkaufen* constitutes a telic verb, *shoppen* is inherently atelic although a transitive, telic use is not generally precluded (e.g. *Klamotten*, *Frische shoppen*). From a connotative point of view, *shoppen* is associated with a pleasant activity, whereas *einkaufen* is emotionally neutral. Computerlinguistic evidence for this contrast, which was anticipated by Kettemann (2006) on the basis of a small set of data, comes from automatically identified co-occurrence partners like *gern(e)*, *Freundin(nen)*, or *Spaß*.

5. Conclusion

In this article, the anglicisms *Dealer*, *Drink*, *Song*, *Skyline* and *shoppen* were compared to semantically close German equivalents. The case studies have shown that contextual information from very large and heterogeneous corpora is a reliable basis for information retrieval and for the performance of contrastive analyses. Although we cannot expect co-occurrence matrices to provide full-fledged semantic representations or encyclopaedic details, prototypical properties of the key words under examination are made accessible to some degree. In order to avoid arbitrariness, the selection of semantically relevant co-occurrences was guided by Pustejovsky's (1996) Generative Lexicon whose multi-dimensional qualia structures provide basic parameters for generic knowledge. These parameters account not only for contrasts between anglicisms and native near-synonyms, but also for instances of semantic approximation.

The automatically performed co-occurrence analyses also revealed lexical properties that are much more difficult to capture out of context, namely connotations, which comprise stylistic, emotional and communicative-pragmatic facets of key words as well as images we typically associate with them. Although connotative nuances are not part of semantic representations, they enrich our knowledge about concepts and allow us to work out very subtle contrasts between loanwords and semantically close native equivalents.

On the whole, the contrastive analyses have shown that despite semantic overlap, the distribution of the anglicisms under consideration and their German near-synonyms is non-arbitrary and largely predictable from their co-occurrence profiles. While *Dealer*, *Drink*, *Skyline* and *shoppen* preserved the specific meaning components with which they were borrowed from English and unfold connotations that are not conveyed by the native near-synonyms *Händler*, *Getränk*, *Stadtsilhouette* and *einkaufen* respectively,

semantic approximation is most obvious for the pair *Lied* and *Song*, the latter of which abandoned its political, socio-critical flavour. Nevertheless, these nouns do not constitute complete synonyms either because *Song* gained the status of a (pseudo-)technical vogue word which distinguishes it from *Lied* in its basic and technical reading. As predicted by Weisgerber (1962: 167), there are no instances of complete synonymy in the examples selected from KANG.

References

Androutsopoulos, Jannis. 1998. *Deutsche Jugendsprache. Untersuchungen zu ihren Strukturen und Funktionen*. Frankfurt a. M. [etc.]: Peter Lang.

Anick, Peter & Pustejovsky, James. 1990. An application of lexical semantics to knowledge acquisition from corpora. In Karlgren, Hans (ed.), *Proceedings of the 13th International Conference on Computational Linguistics*, 7–12, University of Helsinki, Finland.

Baeskow, Heike (2018): "-*Drink* or -*trunk*, that is the question: The distribution of selected anglicisms and German near-synonyms in compounding". *Studia Linguistica* (Online publication ahead of print; https://doi.org/10.1111/stul.12086).

Baroni, Marco & Lenci, Alessandro. 2008. Concepts and properties in word spaces. In Lenci, Alessandro (ed.), From context to meaning. Distributional models of the lexicon in linguistics and cognitive science. *Special Issue of the Italian Journal of Linguistics* 20(1). 55–88.

Belica, Cyril. 2011. Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In Abel, Andrea & Zanin, Renata (eds.), *Korpora in Lehre und Forschung*, 155–178. Bozen: Bozen-Bolzano University Press. Freie Universität Bozen-Bolzano.

Bernadini, Silvia & Baroni, Marco (eds.). 2006. *Wacky! Working papers on the Web as corpus*. Bologna: GEDIT. Available online at http://wackybook.sslmit.unibo.it/

Bouillon, Pierrette, Jezek, Elisabetta, Melloni, Chiara & Picton, Aurélie. 2012. Annotating qualia relations in Italian and French complex nominals". In Calzolari, Nicoletta, Chukri, Khalid, Declerck, Thierry, Uğur Doğan, Mehmet, Maegaard, Bente, Mariani, Joseph, Moreno, Asuncion, Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, 1527–1532. Istanbul: European Languages Resources Association (Elra).

Carlson, Gregory Norman. 1977. *Reference to Kinds in English*. Massachusetts, Amherst: University of Massachusetts. (Doctoral dissertation).

Carstensen, Broder. 1965. *Englische Einflüsse auf die deutsche Sprache nach 1945*. Heidelberg: Carl Winter Universitätsverlag.

Carstensen, Broder & Busse, Ulrich (eds.): *Anglizismen-Wörterbuch. Der Einfluß des Englischen auf den deutschen Wortschatz nach 1945*. Berlin, New York: De Gruyter. Volume 1 (1993), Volume 2 (1994), Volume 3 (1996); reprinted 2001.

Cimiano, Philipp & Wenderoth, Johanna. 2005. Automatically learning qualia structures from the Web. *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, 28–37. Ann Arbor: Association for Computational Linguistics.

Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at http://www.corpusdelespanol.org.

Donalies, Elke 1992. Hippes Hopping und toughe Trendies: Über neumodische, noch nicht kodifizierte Anglizismen in deutschsprachigen Female-Yuppie-Zeitschriften". *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation* 20 (2). 97–110.

Erdmann, Karl-Otto. 1900. *Die Bedeutung des Wortes*. Leipzig: Eduard Avenarius.

Firth, John R. 1957. Modes of meaning. In Firth, John R. (ed.), *Papers in Linguistics* 1934–1951, 180–215. London: Oxford University Press.

Fries, Norbert. 2007. Die Kodierung von Emotionen in Texten. Teil 1: Grundlagen. *Journal of Literary Theory* 1(2). 293–337.

Galinsky, Hans 1963. Stylistic aspects of borrowing. A stylistic and comparative view of American elements in Modern German and British English. In Fraenkel, Ernst, Galinsky, Hans, Gerhard, Dietrich & Lang, H.J. (eds.), *Jahrbuch für Amerikastudien*, 98–135. Band 8. Heidelberg: Universitätsverlag Winter.

Gévaudan, Paul. 2002. *Klassifikation lexikalischer Entwicklungen. Semantische, morphologische und stratische Filiation.*
[http://homepages.uni-tuebingen.de/paul.gevaudan/Filiation.pdf]

Götzeler, Christiane. 2008. *Anglizismen in der Pressesprache. Alte und neue Bundesländer im Vergleich*. Bremen: Hempen Verlag.

Jones, Susan & Sinclair, John M. 1974. English lexical collocations. *Cahiers de Lexicologie* 24. 15–61.

Kettemann, Bernhard. 2006. Morphologische Integration und semantische Differenzierung durch Anglizismen im Deutschen. In Mair, C. (ed.), *Corpora and the history of English. Papers dedicated to Manfred Markus on the occasion of his 65th birthday*, 169–182. Heidelberg: Winter.

Leutloff, Kristina. 2003. Anglizismen in deutschen Jugendzeitschriften. Zu ihren stilistischen und kommunikativen Funktionen am Beispiel von *Bravo* und *Yam*! *Lebende Sprachen* 48. 5–13.

Lipka, Leonhard. 2002. *English lexicology*. Tübingen: Gunter Narr Verlag.

Ludwig, Klaus-Dieter. 1986. Nicht-denotative Informationen lexikalischer Einheiten als Wörterbucheinträge. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39(2). 182–194.

Ludwig, Klaus-Dieter. 2002. Registerkonzepte: Ein Überblick. In Cruse, D. Alan, Hundsnurscher, Franz, Job, Michael & Lutzeier, Peter Rolf (eds.) *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. 1. Halbband, 784–793. Berlin [etc.]: de Gruyter.

Moravcsik, Julius. 1975. Aitia as generative factor in Aristotle's philosophy. *Dialogue* 14. 622–636.

Moravcsik, Julius. 1981. How do words get their meanings? *Journal of Philosophy* 78. 5–24.

Onysko, Alexander. 2007. *Anglicisms in German. Borrowing, lexical productivity, and written codeswitching*. Berlin [etc.]: Walter de Gruyter.

Onysko, Alexander & Winter-Froemel, Esme. 2011. Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43. 1550–1567.

Oxford English Dictionary. http://dictionary.oed.com/

Pfitzner, Jürgen. 1978. *Der Anglizismus im Deutschen. Ein Beitrag zur Bestimmung seiner stilistischen Funktion in der heutigen Presse*. Stuttgart: J.B. Metzlersche Verlagsbuchhandlung.

Poesio, Massimo & Almuhareb, Abdulrahman. 2008. Extracting concept descriptions from the Web: the importance of attributes and values. In Buitelaar, Paul & Cimiano, Philipp (eds.), *Proceedings of the 2008 conference on ontology learning and population: Bridging the gap between text and knowledge*, 29–44. Amsterdam: IOS Press.

Pustejovsky, James. 1996. *The Generative Lexicon*. Cambridge, MA: The MIT Press.

Sosnizka, Julia. 2014. *The collocational behavior of anglicisms in German and American business and news magazines*. Göttingen: Cuvillier Verlag.

Stiven, Agnes B. 1936. *Englands Einfluß auf den deutschen Wortschatz*. Doctoral dissertation, Marburg.

Weisgerber, Leo. 1962. *Von den Kräften der deutschen Sprache. Band I: Grundzüge der inhaltbezogenen Grammatik*. 3. Auflage. Düsseldorf: Pädagogischer Verlag Schwann.

Williams, Edwin. 1981. Argument structure and morphology. *Linguistic Review* 1. 81–114.

Winter-Froemel, Esme. 2011. *Entlehnung in der Kommunikation und im Sprachwandel: Theorie und Analysen zum Französischen.* Berlin/Boston: de Gruyter.

Yang, Wenliang. 1990. *Anglizismen im Deutschen: am Beispiel des Nachrichtenmagazins Der Spiegel.* Tübingen: Max Niemeyer Verlag.

Heike Baeskow
Bergische Universität Wuppertal
Fakultät für Geistes- und Kulturwissenschaften
Anglistik/Amerikanistik
Gaußstraße 20
42119 Wuppertal
Germany
baeskow@uni-wuppertal.de

Jürgen Rolshoven
Universität zu Köln
Institut für Linguistik
Abteilung Sprachliche Informationsverarbeitung
Albertus-Magnus-Platz
50931 Köln
Germany
rols@spinfo.uni-koeln.de

Rules or Constraints? A Cross Methodological Comparison of Approaches to 'Syncope' and 'Vowel Shortening' in Cairene Arabic

Mousa A. Btoosh, Fahad Bin Sultan University, KSA

This study cross-compares three basic phonological approaches to syncope and vowel shortening in Cairene Arabic. The phenomena addressed have been examined within the rule-based, autosegmental and Optimality Theory frameworks with the aim of assessing the effectiveness and shortcomings of using inviolable rules vis-à-vis universal violable constraints. Findings show that the strict application of the general rules and principles within the rule-based and autosegmental approaches often results in some ill-formed output. This, therefore, rationalizes the frequent resort to language-specific rules to account for particular grammatical constructions. Optimality Theory, on the other hand, compels no language-specific restrictions on the input as the optimal form is not emanated by principles or parameters. Rather, it is derived by satisfying the maximum number or incurring the least violations of the relevant universal constraints. Accordingly, principles and generalizations are expressed more straightforwardly and economically with the constraint-based approach of Optimality Theory.

Keywords: *Cairene Arabic, constraints, phonology, syncope, vowel shortening*

1. Introduction

Along with the well-formedness notion, the phonological theory is fundamentally devoted to adequately describing the mental representation and basic units of human speech sounds and explaining how the sound patterns of all languages are underlyingly the same and superficially different (cf. Baković 2013). Consequently, much of phonological research and theorizing during the past five decades has focused on phonological universals and the transition from the area of divergence at the segmental level to the area of convergence at the higher-order prosodic units such as syllables and feet.

A glance at literature on syllable shows the evolution of numerous theories and frameworks since the launch of generative phonology by Chomsky & Halle in 1968. Despite the absence of any significant role for syllable in *The Sound Pattern of English*, the subsequent literature on phonology clearly indicates the central role of syllable in the phonological theory, in general, and phonotactics, in particular. This explains the emergence of several theories primarily dedicated to account for syllable structure such as the skeletal tier theory, the templatic theory and the moraic theory.

The theory of skeletal tier is divided into two models, viz., the CV model and the X-slot model. The CV tier, which was originally proposed by McCarthy (1979), is depicted in (1).

(1) CV Theory

```
   C  V          C V V          C V C
   |  |          | |/           | | |
   t  a          t a           t a t
```
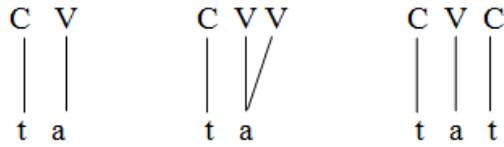
However, the CV model, which does not account for several phenomena, including V-lengthening, was challenged by the X-slot theory. Levin (1985) and Lowenstamm & Kaye (1986) replaced the symbols C and V with a uniform sequence of Xs, as shown in (2).

(2) X-Slot Theory

```
       σ               σ                σ
      / |             / |              / |
     O  R            O  R             O  R
     |  |            |  |\            |  |\
     |  N            |  N C           |  N C
     |  |            |  |\            |  | |
     X  X            X  X X           X  X X
     |  |            |   \/           |  | |
     t  a            t   a:           t a t
```
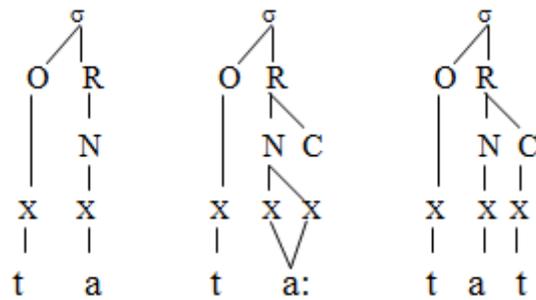
Two main proposals were suggested with regard to the syllable internal structure, namely the level and branching syllable structures. According to the level syllable model, the syllable has a flat structure. The onset comprises the consonant(s) to the left of the nucleus while the nucleus comprises the peak of the syllable. The coda, on the other hand, is made up of the consonant(s) to the right of the nucleus.

(3)

```
              Syllable
             /   |   \
         Onset Nucleus Coda
```
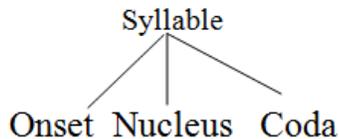
Unlike the level syllable structure, the branching syllable structure has gained much attention in the literature. Two branching syllable structures were, however, proposed in the literature viz., the onset-rhyme model and the body-structure model. The onset-rhyme structure, which gained the most support in the recent literature, is depicted in (4).

(4)

Syllable
Onset Rhyme
Nucleus Coda
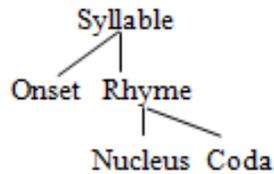
This model was proposed and adopted by Selkrik (1978, 1982) and Halle & Vergnaud (1980) as a linguistic universal drawing on the observed phonotactic constraints that hold between nucleus and coda. The body-syllable structure, which was adopted by several pioneering linguists, including McCarthy (1979) and Iverson & Wheeler (1989), is represented in (5).

(5)

Syllable
Body   Coda
Onset   Nucleus
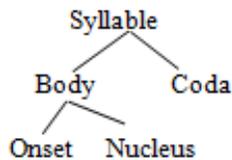
The association between the nucleus and the onset in this model aroused considerable disagreement over the syllable weight as it was noticed in the subsequent literature that the onset does not contribute to the syllable weight (Hyman 1985; McCarthy & Prince 1995; and Hayes 1989). Such debate obviously hastened the emergence of the moraic theory.

Associating segments with weight constitutes the most profound difference between the skeletal tier theory and the moraic theory. Instead of dividing syllables into onset and rhyme under the umbrella of the X-slot theory, syllables are divided into mora or weight units. According to this theory, a light syllable consists of one mora while a heavy syllable is at least bimoraic. The brackets around the mora of the coda in (6) indicate that the weight of the coda (moraic or non-moraic) is context-dependent (weight-by-position) based on a language specific basis (Topintzi 2011).

(6)

σ
μ (μ)
C V C
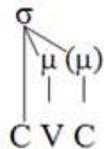
A survey of literature also shows some disagreement as to where the onset associates and whether or not it is extramoraic. For Hayes (1989), the onset is non-moraic and directly adjoins to the

syllable node as in (6) whereas it attaches to (and shares the same mora with) the following nucleus, as shown in Figure (7) (Hyman 1985; Itô 1986, to name but a few). Support for the existence of onset weight also comes from Topintzi (2006, 2010).

(7)
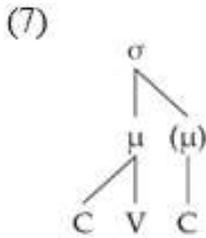
Nevertheless, there are some differences between rhyme-weight and nucleus-weight languages, as depicted in (8a) and (8b), respectively (Ewen & van der Hulst 2001: 151).

(8)
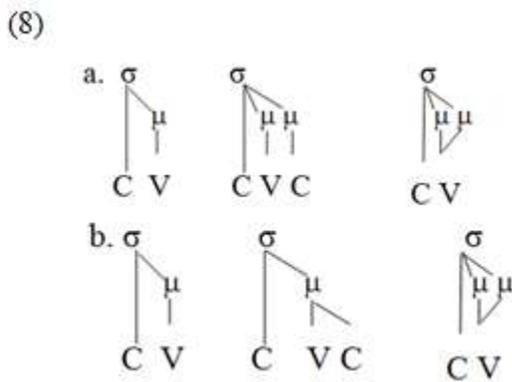
Two significant findings emerge from the figures in (8). First, the initial consonants are always extramoraic. Also, unlike rhyme-weight languages, the final consonant in nucleus-weight languages does not associate with a separate mora. The marked variations across languages and dialects and the inability of the previous approach to adapt to these dissimilarities were the chief reason underlying the need for a new model of grammatical representation that is capable to accommodate such variations (Vihman & Croft 2007). This rationalized the emergence of the template-based approach.

The templatic theory of syllabification proposed by Itô (1986, 1989) attempts to integrate "well-formedness conditions on syllable structure into syllabification algorithm built around directional template matching" (Frampton 2008: 228). Syllabification, based on this theory, proceeds by mapping segmental units onto a given template. It is still crucial to point out that syllable structure is assigned in one single direction. So, a consonant that does not fit the template condition is prosodically unrealized or unlicensed (Martínez-Gil 1991: 554). As such, unlicensed segments are either deleted or rescued via vowel epenthesis. However, the systematic and significant exceptions to rule-based theories, which are always accounted for by proposing

language specific rules, provided a fertile ground for the evolvement of the constraint-based approach of Optimality Theory (OT, henceforth).

Unlike rule-based and autosegmental theories, syllable well-formedness within the OT framework is governed by a set of universal constraints. Language specific rules are, however, accounted for by different hierarchical ranking of constraints. In OT, the relation between the phonemic form and the phonetic form is mediated by two formal mechanisms called *GEN(erator)* and *EVAL(uator)*. GEN, on the one hand, creates a set of output candidates that will undergo the comparison and judgment of the EVAL. All candidates are simultaneously evaluated for whether or not they are in accord with the set of constraints included in the tables. Since OT depends on comparison, then the candidate's first violation(s) cannot tell whether it is optimal or not. However, the output is chosen from the set of candidates created by the GEN (regardless the number of such candidates) (Prince & Smolensky 1993). Languages' adherence to universal constraints is never absolute and variations among varieties can be accounted for not by positing new or different rules as was the case under the umbrella of earlier models, but rather by proposing a hierarchical system of both violable and ranked constraints (e.g. Prince & Smolensky 1993; Tranel 1995; Btoosh 2006).

This study attempts to shed light on syncope and vowel shortening in CA within the framework of the traditional approach, X-slot theory, moraic theory and OT. To gain a transparent picture of the target phonological phenomena, a deliberate attempt has been made to ensure that the same examples are used throughout the study. Moreover, in conformity with the aims posited above, close attention has been paid to how each approach accounts for the exceptions to the general rules with the aim of delineating the best approach in terms of ease, consistency, economy and highest conformity to general, invariant principles cross-linguistically.

2. Syllable structure in CA

Three syllable patterns surface in CA: light CV, heavy CVC and CVV, and superheavy CVVC and CVCC. It is noteworthy that the final C in CVC, CVCC and CVVC is extrasyllabic or invisible word finally. It is also worth mentioning that superheavy syllables occur only in word or phrase-final position (Aquil 2013).

Given the syllable patterns shown above, it is evident that onsetless syllables are banned from surfacing in this variety. Likewise, consonant clusters are banned word-initially or medially. So, in order to avoid having onsetless syllables at word level, CA resorts to glottal stop epenthesis or resyllabification to fill in the empty slot. CA also turns to resyllabification to avoid having consonant clusters across word-boundaries since utterance is the syllabification domain in Arabic, in general, as shown in (9).

(9)

	UR		SR	Process
a.	akbar	→	?ak.bar	onset-motivated epenthesis
	'bigger'			

 b. kalb+ak →   kal.bak   onset-motivated resyllabification
   dog your
   'your dog'

 c. katabt-   l- ha→ ka.tab.til.ha phrasal resyllabification
   wrote-1sg.mas. to her
   'I wrote to her'

3. Syncope and vowel shortening in CA

Like other Arabic varieties, CA exhibits syncope and vowel length alternation phenomena. Syncope in CA refers to a general phonological process, whereby a high vowel deletes in a non-final monomoraic syllable that is flanked by vowel-final syllables across words (Watson (2007). However, it should be made obvious that syncope in all Arabic dialects that do not allow complex margins (such as CA) is not context-free. Rather, the syncopated vowel should be preceded by an open syllable (Kenstowicz 1980; Watson 2002, to name but a few).

 (10)
  a. fih**i**m   l- gawaab → fih.mil.ga.waab
   understood 3sg. mas. the answer
   'he understood the answer'

  b. ʃir**i**b   l- ʕasiir   → ʃir.bil.ʕa.siir
   drank 3sg.mas. the juice
   'he drank the juice'

Syncope in CA also takes place within a phonological word when a vowel-initial suffix is added.

 (11)

  a. libis+u →   lib.su
   dressed 3pl.mas.
   'they dressed up'

  b. ʃir**i**b+u →   ʃir.bu
   drank 3pl.mas.
   'they drank'

Despite the attested deletion of the short vowels in non-final monomoraic syllables, it is yet quite noticeable that the syncopation of high short vowels is more common than low short ones. A close look at literature of Arabic varieties shows a distinction between differential and nondifferential

dialects. Differential dialects are those varieties that delete unstressed high vowels while nondifferential ones delete all short vowels whether high or low (Cantineau 1939). CA belongs to the first group (differential dialects). Unlike syncope, which rules out marked but not ill-formed structures, vowel shortening is utilized in CA to prevent ill-formed syllables from surfacing.

Vowel shortening refers to the reduction of long vowels in closed non-final syllables. Research on Arabic has demonstrated the existence of both phenomena in Arabic varieties (McCarthy 1979; Kenstowicz & Abdulkareem 1980; Abu Salim 1982; Irshid 1984; Alghazo 1987; Abu Mansour 1995; Farawneh 1995; Watson 2007, among others). It is yet noteworthy that vowel shortening and syncope usually coexist in the same grammar so as to improve the foot shape (Gouskova 2003: 86). Further evidence in favor of this argument is provided by examples such as the ones in (12).

(12)

    a. naadim + iin ⬚ nad.miin
        regretful 3pl.mas.
        'they are regretful'

    b. ṣaaħib+u ⬚ ṣaħ.bu
        friend his
        'his friend'

3.1 Syncope and vowel shortening in CA: A cross comparison of approaches

3.1.1 A rule-based approach

Irrespective of whether syncope is triggered by a markedness constraint that disfavors weak nucleus syllables (Mobaidin 1999; Gouskova 2003, among others) or by economy so as to reduce the number of monomoraic syllables, it is still essential to point out that this process is not constraint free, as formalized in (13).

(13)

$$\left\{\begin{bmatrix} \text{V} \\ \text{+ high} \\ \text{- stress} \end{bmatrix}\right\} \longrightarrow \varnothing \ / \ \text{VC} \ \underline{\quad} \ \text{CV}$$

Building on this rule, only unstressed high short vowels are subject to deletion if they occur in a nonbranching rhyme preceded by an open syllable. In context of stress, it is essential to mention that stress in CA is predictable and governed by weight, as shown in (14).

(14) Word-stress patterns in CA

    a. A final superheavy (CVCC or CVVC) syllable is stressed.

i. ?a'kalt
  ate 1sg.mas.
  'I ate'

ii. miħ'raab
  'niche'

    b. If the last syllable is not superheavy, stress the penultimate heavy syllable.

      i. ʃi'ribti
        drank 2sg.fem. (intonational question)
        'did you drink?'

      ii. iʃta?'t-illak
        longed 1sg.mas.-for you
        'I longed for you'

    c. Otherwise, stress falls on either the penult or the antepenult whichever is separated from a preceding heavy syllable (or word boundary) by an even number of light syllables, including zero".

      i. ?in'katabat
        it (fem.) was written
        'it was written'

      ii. 'kataba
        wrote 3sg.mas.
        'he wrote'

However, stress is assigned to the penultimate syllable in the third person feminine singular inflectional form of the perfective syllable with a V(C) object suffix and in a plural with the template CiCiCa or CuCuCa (Watson 2007).

(15)
      i. kata'bitu
        wrote 3sg.fem. it
        'she wrote it'

ii.   ɣiˈriba
       'crows'

The examples in (16) below show that syncope takes place at word level when a vowel-initial suffix is added. It should be, however, made obvious that syncope must be averted if it leads to impermissible structures, as shown in the examples sketched in (16a).

(16)
a. Syncope Inhibited   b. (Input) →   (output) Syncope Obligatory
  i.   fihim   i. fihim+u   →   fih.mu
        understood 3sg.mas.   understood 3pl.ma:
        'he understood.'   'they understood'

  ii.   bitiktib   ii. ʃufit+ak   →   ʃuf.tak
        write 3sg.fem.   saw 2sg.mas.
        'she is writing/writes'   'I saw you'

  iii. ʔuxtina   iii. biyakil+u   →   bi.yak.lu
        sister our   eat 3pl.mas.
        'our sister'   'they are eating'

Note that it is banned to delete the high short [i] in (16ai) 'fihim' since this leads to a violation of the scale of sonority as [m] is more sonorous than [h]. So, it is impossible for [m] to be syllabified as a coda of the preceding syllable. Also, it is disallowed to syllabify the [h] as the coda of the first syllable since this renders the second syllable onsetless. Over and above, the deletion of the unstressed high short vowel in (16a ii & iii) is inadmissible as this leads to a non-permissible structure.
       The following examples show that syncope also applies at phrase level due to resyllabification across word boundaries.

(17)

    i. simiʕ   +   l- kalaam   →   sim.ʕil.ka.laam
       obeyed 3sg.mas. the advice
       'he obeyed the instructions/advice'

    ii. naam +   kiθiir   →   naa.mik. θiir
        slept 3sg.mas.   much
        'he slept much"

67

A closer look at the data above shows that CA, which does not allow CCC sequence, prefers vowel insertion rather than segment deletion to rescue the unlicensed segments. Epenthesis in CA can be expressed in the following rule.

(18)

$$\emptyset \longrightarrow \left\{ \begin{bmatrix} V \\ +high \\ -stress \end{bmatrix} \right\} / \ C.C\text{-}C$$

Despite the seeming complementarity of syncope and epenthesis, they are plainly triggered by different reasons. While epenthesis breaks up a marked consonant cluster, syncope, which deletes a medial, typically unaccented vowel, is induced by economy. Nevertheless, syncope often occurs as a result of epenthesis, as shown below.

Epenthesis, which projects a mora to the right of an unstressed syllable, is, thus, a repair operation since the pre-epenthesis form does not conform to the basic syllable structure of CA. A look at the resyllabification of the examples above displays the interaction between epenthesis and syncope as the first feeds the second. Nevertheless, epenthesis is not the only repair strategy in CA. Rather, most Arabic varieties, including CA, employ vowel shortening to avoid impermissible syllable structures.

Shortening of input long vowels in non-final syllables is a must in all CV dialects, most varieties spoken in Egypt and the Libyan desert, including CA, since these dialects do not license three mora syllables (Kiparsky 2003). Thus, CVVC syllables are eliminated by shortening the input vowel, as formalized in (19).

(19)

$$\left\{ V: \right\} \longrightarrow V / \_\_ \ C]_\sigma$$

In the examples presented below, it is easy to notice not only the rules applied to the underlying form, but also the interactions among such rules. For instance, the epenthetic vowel sometimes creates an environment that makes vowel syncopation possible, as in (20a). Therefore, accounting for the structural change presented above requires rule ordering due to the obvious feeding relation among such rules.

(20)

a. # biʕt+kitaab#	b. #kitaab#	c. #kitaab+ha#	d. #ṣaaħib+u#	UR
sold1mas.sg.+ book	*book*	*book her*	*friend his*	
'I sold a book'	*'book'*	*'her book'*	*'his friend'*	
biʕ.ti.ki.taab	N/A	N/A	N/A	Epenthesis
biʕ.tik.taab	N/A	N/A	ṣaaħ-bu	Syncope
N/A	N/A	kitab-ha	ṣaħ-bu	Shortening
biʕ.tik.taab	kitaab	kitab-ha	ṣaħ-bu	SR

Had the rule ordering of epenthesis and syncope been reversed, syncope would not have been applied, as shown in (20). Therefore, vowel syncopation in (20a) applies to the output of the epenthesis rule. The same is true for syncope and vowel shortening, as in (20d). Overall, the examples in (20) empirically demonstrate that epenthesis, syncope and vowel shortening aim to respectively insert, delete and reduce segments and structures that would be otherwise impermissible in this variety.

3.1.2 *An x-slot account*

On the basis of resemblance between the CV and X-Slot models as demonstrated above, it is evident that the latter is an improved version of the former. The x-slot model easily accounts for phonological processes related to timing such as compensatory lengthening. That is, each x-slot represents a timing unit. Therefore, a short sound occupies one x-slot in the rhyme while a long sound occupies two x-slots. Yet, this theory treats onsets and nuclei/codas equally with regard to weight (Elfner 2006). Like other theories that endorse the existence a syllabic skeleton, elements of the segmental tier are associated with elements of the CV-tier in accordance with Universal Association Convention. Associations are also constrained by the Well-formedness Principle (Goldsmith 1976; Watson 2007). Accordingly, crossing association lines are banned. Likewise, association of consonantal elements on the segmental tier with vowels on the CV-tier or vice versa is not allowed. Yet, having an element on one tier associated with two or more elements on the other remains possible (Berendsen 1986). The basic syllable structure is represented in the following figures.

(21) Associate each vowel with a syllable nucleus.
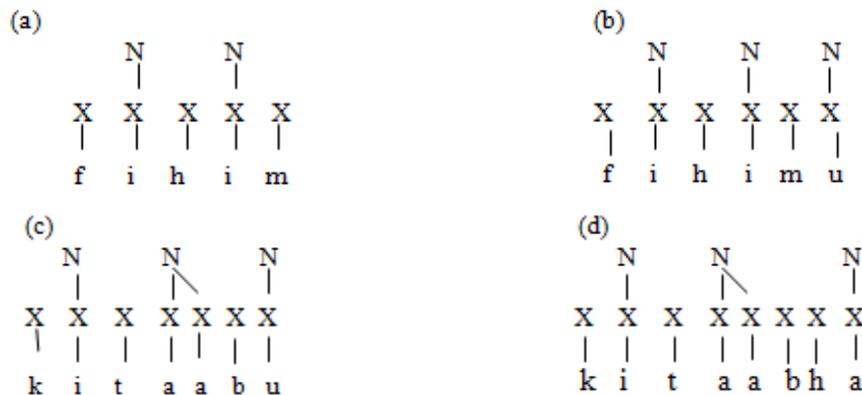
The vocalic slot in each syllable is pre-associated to a nucleus. Also, the examples above show that the vowel element in light syllables is associated with one X in the rhyme while a heavy syllable is associated with two Xs. Next to vowel association, consonants to the left of the nucleus are associated to the onset node.

(22)

    (i) Onset Syllabification: Associate the segments to the left of the nucleus with an onset.

(ii) Onset Maximization: Maximize the complexity of the onset, subject to relevant constraints.

(a)
```
O   N   O   N
|   |   |   |
X   X   X   X   X
|   |   |   |   |
f   i   h   i   m
```

(b)
```
O   N   O   N   O   N
|   |   |   |   |   |
X   X   X   X   X   X
|   |   |   |   |   |
f   i   h   i   m   u
```

(c)
```
O   N   O   N       O   N
|   |   |   /\      |   |
X   X   X  X X     X   X
|   |   |   \/      |   |
k   i   t   a:      b   u
```

(d)
```
O   N   O   N           O   N
|   |   |   /\          |   |
X   X   X  X X     X   X X
|   |   |   \/          |   |
k   i   t   a:      b   h   a
```
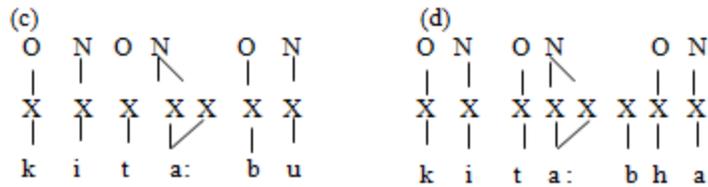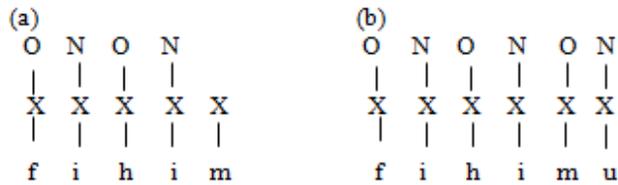
(23) Coda Syllabification: Associate the segments to the right of the nucleus with a coda, subject to relevant constraints.

(a)
```
O   N   O   N   C
|   |   |   |   |
X   X   X   X   X
|   |   |   |   |
f   i   h   i   m
```

(b)
```
O   N   O   N   O   N
|   |   |   |   |   |
X   X   X   X   X   X
|   |   |   |   |   |
f   i   h   i   m   u
```

(c)
```
O   N   O   N       O   N
|   |   |   /\      |   |
X   X   X  X X     X   X
|   |   |   \/      |   |
k   i   t   a:      b   u
```

(d)
```
O   N   O   N           C   O   N
|   |   |   /\          |   |   |
X   X   X  X X     X   X X
|   |   |   \/          |   |   |
k   i   t   a:      b   h   a
```
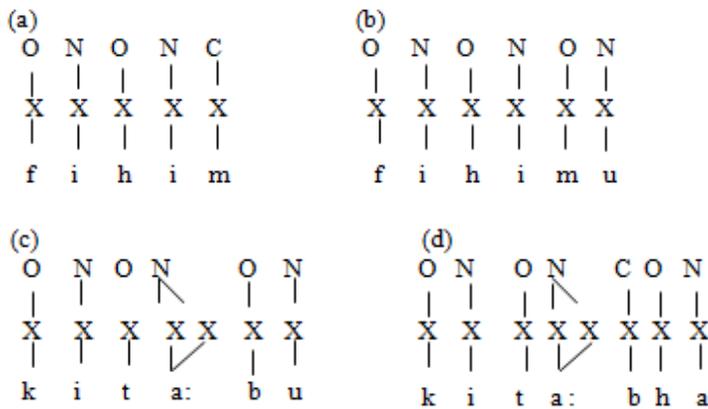
Thus far in this section, an attempt has been made to apply the universal syllabification rules. However, the strict application of such rules has resulted in some ill-formed output forms, as shown in (24b) and (24d) below.

(24)
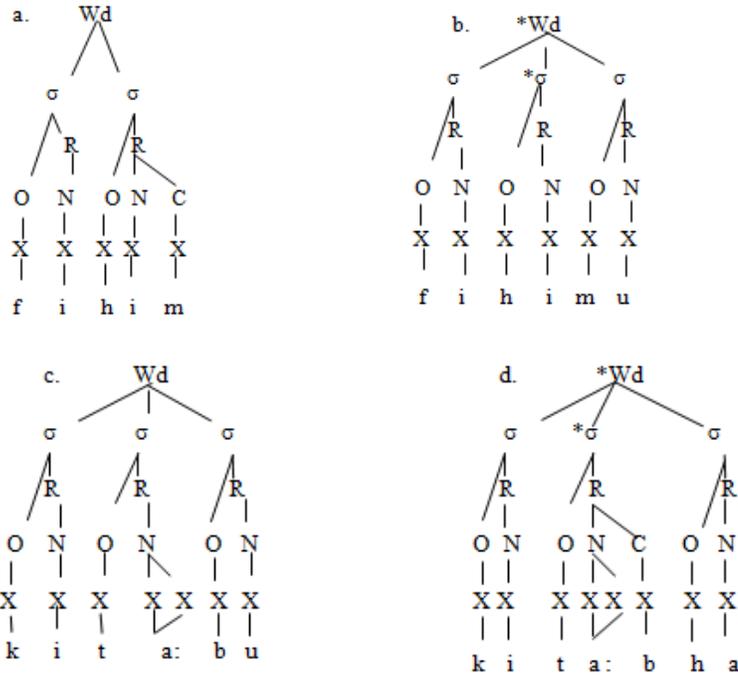
The ill-formedness of (24b) is attributed to the surfacing of the unstressed high short [i] in a nonbranching rhyme preceded by an open syllable. This is definitely not licensed in this variety. As shown in (25a), the ill-formed syllabification of (24b) is rescued by deleting the [i] and associating the [h] to the coda of the preceding syllable. In (24d), however, it is impermissible for CVVC to occur word medially. It is also disallowed to associate the [b] to the onset of the ultimate syllable as this results in an inadmissible syllable: *CCV.

Consequently, in order to avoid having the ill-formed output forms in (24b) and (24d) above, CA resorts to two phonological processes, namely, vowel syncopation as in (25a) and vowel shortening as in (25b). Therefore, syncopation is meant to avoid having a nonbranching rhyme word-medially while vowel shortening serves to avoid having extraheavy syllable word-initially or medially.

(25)

71
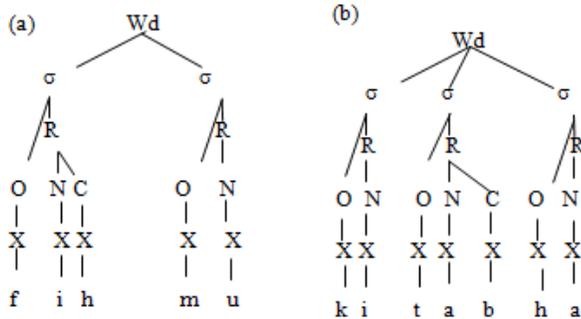
3.1.3 *A moraic analysis*
Unlike the previous syllable models, the moraic theory, which replaces the hierarchical structure of the X-slot skeleton with moras, states that syllable weight functions as a phonological variable. Weight, according to this theory, may vary from one language to another. However, a uniform weight criterion is employed within a given language. Nuclei are weight units or mora positions while onsets are non-moraic. Codas, on the other hand, are not underlyingly moraic. Rather, they become weight units by the Weight-by-Position rule (Hayes 1989). This provides an explanation as to why CVC syllables are considered heavy in certain languages and light in others.

As far as CA is concerned, CV syllables are always light while CVC syllables are only light in the domain-final position. That is, the last C in CVC syllables is always extrametrical in Arabic varieties. Thus far, it is apparent that the moraic theory imposes a syllabification algorithm, which determines how individual segments are to be parsed into a syllable. Moreover, it provides a straightforward account for the issue of segment quantity (how long and short segments are assigned different configurations). Nonetheless, applying the syllabification rules to the input data sometimes results in ill-formed output, as shown in (26).

(26)
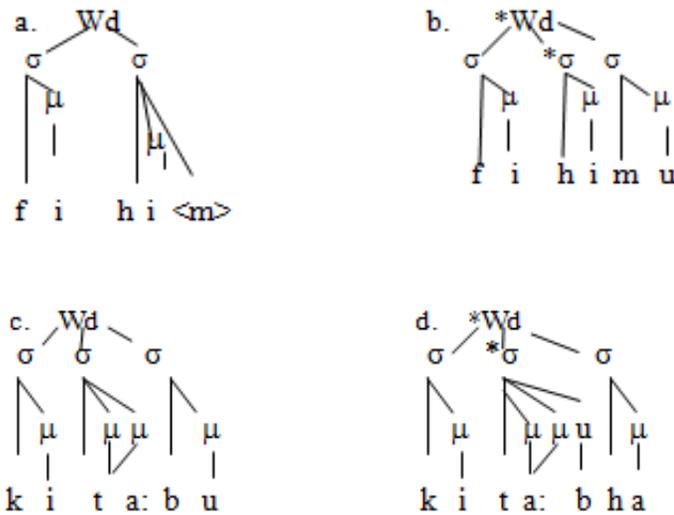
The ill-formedness of (26b) is attributed to a language specific rule which bans the surfacing of a monomoraic syllable (word-medially) if preceded by another monomoraic syllable. For this purpose, Arabic varieties employ vowel syncopation not to repair an ill-formed structure but rather to minimize monomoraic syllables, as depicted in (27).

(27)
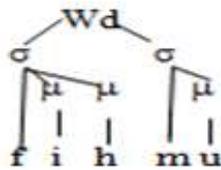
Unlike (26b), the ill-formedness of (26d) is ascribed to the surfacing of a trimoraic syllable word-medially. As was illustrated above, CVXC syllables are restricted to word-final position in this variety. To this end, CA resorts to vowel shortening to avoid having a trimoraic syllable word-medially.

(28)
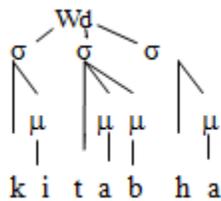
Drawing on the above analysis, syncopating vowels and vowel length alternations are meant to produce syllables that are pronounceable in the language. To be more exact, CA allows for no monomoraic syllables word-medially and no trimoraic syllables word-initially or medially. In the context of superheavy syllables, it should also be noted that the domain-final consonants in superheavy syllables (CVXC) are prosodically licensed, but as extrasyllabic (Watson 2007; Aoun 1979; Hayes 1979, among others).

3.1.4 *An optimality-theoretic account*
Syllable structure, within the OT framework, is governed by markedness and faithfulness constraints. Markedness constraints impose conditions on the output well-formedness while faithfulness constraints impose the exact preservation of the input in the output. As far as onset is concerned, CA, as shown below, is in full compliance with the universal unmarked constraint.

(29) Ons

    Every syllable has an onset.   (Prince & Smolensky 1993)

In order to prevent the underlyingly onsetless syllables from surfacing in CA, the variety resorts to either epenthesis, whereby a glottal stop is inserted to fill in this empty slot or resyllabification whereby a coda resyllabifies as an onset of the following onsetless syllable. Consequently, CA prefers adhering to the markedness constraint Ons to breaching the faithfulness Dep-io constraint.

(30) DEP-IO

Every segment of the output has a correspondent in the input (prohibits   phonological epenthesis)   (McCarthy & Prince 1995)

(31) Input: akbar 'bigger'

Table 1: ONS >>DEP-IO

/akbar/	ONS	DEP-IO
a. ak.bar	*!	
b. ☞?ak.bar		*

The first candidate is ruled out as CA strictly bans onsetless syllables. Despite incurring a violation of the DEP-IO, candidate (b) could win by successfully managing to escape the fatal violation of the top-ranked constraint ONS.

As mentioned earlier, CA allows for no monomoraic syllables word-medially. As such, a weak nucleus (*WN) or a high-unstressed short vowel deletes in an open syllable if preceded by an open syllable (Mobaidin 1999; Watson 2002). No doubt, the elision of the weak nucleus results in an inevitable violation of the MAX-V-IO constraint, which prohibits the input vowel deletion. Moreover, it should be reiterated that syncope must not result in impermissible structures, including complex margins.

(32) *WN

A high short vowel in an open unstressed syllable must be deleted if preceded by an open syllable.

(33) MAX-V-IO

Input vowels must have output correspondents. ('No vowel deletion.')

(Kager 1999)

(34) *COMPLEX

Syllables have at most one consonant at edge.

(Archangeli 1997)

(35)   Input: fihim+u
                understood 3pl.mas.
                'they understood'

Table 2: *WN, *COMPLEX >>MAX-V-IO

/fihim+u/	*WN	COMPLEX	MAX-V-IO
a. ☞fih.mu			*
b. fi.hi.mu	*!		
c. fi.hmu		*!	*

Although it is not in harmony with the MAX-V-IO, candidate (a) emerges as the winner by just avoiding the penalty of the top-ranked constraints *WN and *COMPLEX. Candidate (b), on the other hand, is ruled out as it contains a monomoraic syllable word internally. The last candidate is eliminated by the highly ranked *COMPLEX as consonant clusters in this variety are banned except in utterance-final position.

CA, as shown above, allows for no trimoraic syllables word-initially or medially. As a result, it resorts to vowel shortening and extrasyllabicity to avoid having such type of syllables.

(36) *3μ
No trimoraic syllables.
(Kager 1999)

(37) MAX-μ "No shortening"
"For every V that corresponds to V' in the output, every μ that is linked to V has a correspondent μ' linked to V'." (McCarthy & Prince 1995)

(38)   Input: kitaab+ha
                book her
                'her book'

Table 3: *3μ,*COMPLEX >> MAX-μ

/kitaab+ha/	*3μ	*COMPLEX	MAX-μ
a. ki.taab.ha	*!		
b. ☞ki.tab.ha			*
c. ktaab.ha	*!	*	*

Candidate (a) represents a successful attempt to remain faithful to the input form. Yet, it is ruled out as it fatally violates the top-ranked *3μ constraint, which disallows having trimoraic syllables. In spite of not being in harmony with the faithfulness constraint (MAX-μ), the second candidate emerges as the winner by completely satisfying the high-ranked constraints. Candidate (c), the least harmonic among all the competing forms in terms of the number of the incurred violations, is precluded as a consequence of incurring fatal violations against the two top-ranked markedness constraints (*3μ and *COMPLEX).

Syncope also applies to syllables across word boundaries when similar conditions are met, as illustrated in (39).

(39) Input: biʕt + kitaab
            sold 1sg.mas. + book
        'I sold a book'

Table 4: *Wɴ, *Cᴏᴍᴘʟᴇx >> Dᴇᴘ-ɪᴏ, Mᴀx-v-ɪᴏ

/biʕt+kitaab/	*WN	*COMPLEX	DEP-IO	MAX-V-IO
a. biʕt.ki.taab		*!		
b. ☞ biʕ.tik.taab			*	*
c. biʕ.ti.ktaab		*!	*	*

The grammar of the variety penalizes the first candidate and eliminates it from consideration due to its violation of the *Cᴏᴍᴘʟᴇx since branching codas are allowed only word or utterance finally. Though the high short vowel in the second syllable in (b) has no input correspondent, epenthesis remains the only phonological process available to rescue the optimal candidate. Accordingly, the winner has sacrificed the low-ranked constraints (DEP-IO and MAX-V-IO) in favor of the top-ranked ones *WN and *COMPLEX. The last candidate, on the other hand, is ruled out by incurring a fatal violation of the *COMPLEX constraint.

4. Conclusion

Despite the superficial diversity among languages at all levels, languages still share quite fundamental similarities, which are attributed to the existence of the innate principles. Using different linguistic approaches, this study examined syncope and vowel shortening in CA. Drawing on the mechanisms and findings of the approaches employed here, it has become obvious that the application of the generalizations and principles does not guarantee correct output due to language specific rules. This, therefore, explains the frequent resort of these theories and approaches to parameters to account for the language specific rules.
    Based on how such approaches deal with exceptions and parameters, it is evident that only OT is capable of accounting for parameters under the umbrella of the universal principles without any need to proposing any specific constraints. Thus, the answers provided by the OT are more straightforward and economical than the answers provided by any other approach.

References

Abu Mansour, Mahasen Hasan. 1995. Optimality and conspiracy in the syllable structure of Arabic. In Beckman, Jill & Dickey, Laura Walsh & Urbanczyk, Suzanne (eds.), *University of Massachusetts Occasional Papers: Papers in Optimality Theory* 18.1-20.

Abu-Salim, Isam M. 1982. *A Reanalysis of Some Aspects of Arabic Phonology: A Metrical Approach*. Urbana, IL: University of Illinois. (Doctoral dissertation.)

Alghazo, Mohammad H. 1987. *Syncope and Epenthesis in Levantine Arabic: A Nonlinear Approach*. Urbana, IL: University of Illinois. (Doctoral dissertation.)

Aoun, Joseph. 1979. Is the syllable or the supersyllable a constituent? *MIT Working Papers in Linguistics* 1:140-148. Cambridge, Mass: MIT Department of Linguistics and Philosophy, MITWPL.

Aquil, Rajaa. 2013. Cairne Arabic Syllable Structure though Different Phonological Theories. *Open Journal of Modern Linguistics*, 3. 259-267.

Baković, Eric. 2013. Phonology and phonological theory. In Goldrick, Matthew & Ferreira, Victor & Miozzo, Michele (eds.), *The Oxford Handbook of Language Production*, 199-209. Oxford: OUP.

Berendsen, Egon. 1986. *The phonology of cliticizations*. Dordrecht: Foris.

Btoosh, Mousa A. 2006. Constraint interactions in Jordanian Arabic phonotactics: An optimality-theoretic approach. *Journal of Language and Linguistics* 5(2). 192-221.

Cantineau, Jean. 1939. Remarques sur les parlers de sédentaires syro-libanopalestiniens. Bulletin *de la Société de Linguistique de Paris*, 40. 80-88.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Elfner, Emily. 2006. Contrastive syllabification in Blackfoot. In Baumer, D. & Montero, D., & Scanlon, M., (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*. 141–149, Somerville, MA. Cascadilla Proceedings Project.

Ewen, Colin J. & van der Hulst, Harry. 2001. *The Phonological Structure of Words*. Cambridge: Cambridge University Press.

Farwaneh, S. 1995. *Directionality effects in Arabic dialect syllable structure*. Salt Lake City: University of Utah. (Doctoral dissertation.)

Frampton, John. 2008. SPE Extensions: Conditions on Representations and Defect-Driven Rules. In Vaux, Bert & Nevins, Andrew (eds.), *Rules, Constraints and Phonological Phenomena*, 220-251. Oxford: Oxford University Press.

Goldsmith, John. 1976. An overview of autosegmental phonology', *Linguistic Analysis*, 2. 23-68.

Gouskova, Maria 2003. The reduplicative template in Tonkawa. *Phonology* 24. 367-396.

Halle, Morris, & Vergnaud, Jean-Roger. 1980. Three-dimensional phonology. *Journal of Linguistic Research* 1:83–105.

Harris, John & Gussman, Edmund. 2003. *Word-final onsets.* ROA. 575-0203

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20 (2). 253-306.

Hulst, Harry van der & Ritter, Nancy A. 1999. *Head-driven phonology*. In Hulst, Harry van der & Ritter, Nancy (eds.), *The syllable: views & facts*, 113- 167. Berlin: Mouton de Gruyter.

Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris Publications.

Irshid, Omar. 1984. *The phonology of Bani Hasan Arabic*. Urbana, IL: University of Illinois. (Doctoral dissertation.)

Itô, Junko.1986. *Syllable theory in prosodic phonology*. Amherst: UMass. (Doctoral dissertation.)

Itô, Junko. 1989. *A prosodic theory of epenthesis*. NLLT 7. 217-259.

Iverson, Gregory K. & Wheeler, Deirdre W. 1989. Phonological categories and constituents. In *Linguistic Categorization*. Amsterdam: Benjamins.

Kager, René. 1999. *Optimality Theory*. Cambridge: CUP

Kenstowicz, Michael. 1980. Notes on Cairene Arabic syncope. *Studies in the Linguistic Sciences*, 10. 39-53.

Kenstowicz, Michael J. & Abdul-Karim, Kamal. 1980. Cyclic Stress in Levantine Arabic. *Studies in the Linguistic Sciences* 10. 77-98.

Kiparsky, Paul. 2003. Syllables and moras in Arabic. In Féry, Caroline & van de Vijver, Ruben (eds.) *The syllable in Optimality Theory*. 147–182. Cambridge: Cambridge University Press.

Levin, Juliette. 1985. *A metrical theory of syllabicity*. Cambridge: MIT. (Doctoral dissertation.)

Lowenstamm, Jean & Kaye, Jonathan. 1986. Compensatory lengthening in Tiberian Hebrew. In Wetzels, Leo & Sezer , Engin. (eds.), *Studies in Compensatory Lengthening*, 97-132. Dordrecht: Foris.

Martínez-Gil, Fernando. 1991. The insert/delete parameter, redundancy rules, and neutralization processes in Spanish. In Campos, Héctor & Martínez-Gil, Fernando (eds.), *Current Studies in Spanish Linguistics*. 495 - 571.

McCarthy, John. 1979. On stress and syllabification, *Linguistic Inquiry*, 10. 443 - 465

McCarthy, John & Prince, Alan. 1995. Faithfulness and reduplicative identity. In Beckman, Jill & Dickey, Laura Walsh & Urbanczyk, Suzanne, (eds). *University of Massachusetts Occasional Papers: Papers in Optimality Theory* 18, 249-384.

Mobaidin, Hosam Eddin. 1999. Consonant underpaying in Arabic: An optimality theoretic approach. *Dirasat, Proceedings of F.I.C.A.E.C.C.S Special Issue*, University of Jordan.

Prince, Alan & Smolensky, Paul. 1993. *Optimality Theory: Constraint interaction in generative grammar*. RuCCS-TR-2. ROA-537.

Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, Mass. & Oxford: Blackwell.

Selkirk, Elisabeth O. 1978. *On Prosodic structure and Its relation to syntactic structure*, unpublished MS, Amherst: University of Massachusetts.

Selkirk, Elisabeth O. 1982. The syllable. In van der Hulst, Harry & Smith, Norval (eds.), *The Structure of Phonological Representations*, Part II, 337-383. Dordrecht: Foris.

Topintzi, Nina. 2006. *Moraic onsets*. London: University College London. (Doctoral dissertation.)

Topintzi, Nina. 2010. *Onsets: suprasegmental and prosodic behaviour*. Cambridge: Cambridge University Press.

Topintzi, Nina. 2011. Onsets. In van Oostendorp, Marc & Ewen, Colin J. & Hume, Elizabeth & Rice, Keren (eds.), *The* Blackwell *companion to phonology*, 1285–308. Oxford: Blackwell Publishing.

Tranel, Bernard. 1995, *Rules vs. constraints: a case study*, ROA-72.

Vihman, Marilyn & Croft, William. 2007. Phonological development: Towards a "radical" templatic phonology. *Linguistics* 45 (4): 683-725.

Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Watson, Janet C. E. 2007. *The phonology and morphology of Arabic*. Oxford: Oxford university press.

Mousa A. Btoosh, Ph.D.,
Dept. of English Language and Translation
Fahad Bin Sultan University
P.O. Box 15700
Tabuk 71454, KSA
Cell. (966)566 32 3942
e-mail: mbtoosh@fbsu.edu.sa
    mousa.btoosh@gmail.com

Expressing Social Deixis through Prosody:
A Case Study of American English

Magdaléna Bilá, Alena Kačmárová, Marianna Kraviarová

Prešov University

Expressing interpersonal relationships varies from language to language. Numerous studies explore morphological or other means of expressing a tous/vous relationship. To our knowledge, no research has been done on whether the phonic realization has a share in mapping a T or V shade onto an utterance. The present study presents the results of such research. After the corpus was compiled and T and V utterances categorized, we measured pauses and melody contours, and we identified the pitch accent placement. Then, we interpreted the data sociolinguistically. The data point to two areas worth further examination – phonetic and sociological: a) a tendency was observed in T vs. V encounters with regard to the sociological parameter of age; b) the American culture seems to apply the model of "dispersion" rather than bipolarity, which makes it an intricate task to collect a sufficient number of V encounters providing for statistically significant data.

Keywords*: social deixis, T/V forms, temporal/force/tone modulation of a speech signal, ethnomethodological description, sociolinguistic interpretation*

1. Introduction

The verbal vocal realization of social relationships makes available a picture of conventionalized communication styles. Expressing interpersonal relationships in languages necessarily has social and form-related aspects. From a social perspective, it is largely a matter of a sliding scale on which (very) formal through (very) informal expressions are placed. Formally (in terms of the inventory of pronominal or morphosyntactic means), however, in most languages, conveying social deixis is dichotomic, i.e. it oscillates between two poles: T and V forms. This is the case of most European languages (Slavic languages, Romance languages, Germanic languages). Since these languages have in possession and, thus, use explicit pronominal and morphological markers, the users are able to consciously choose proper expression of social distance. The formal dichotomy is reflected overtly through pronominal forms and verb morphology. The presence of specific forms in the lexicon and verb morphology contributed to the conventionalized conscious usage of particular sequences in particular communication acts. Thus, the overtness implies a speaker's knowledge and subsequent usage of specific language means as a component of language behavior in T (*tous*) or V (*vous*) settings, i.e. T/V relationships.

Present-day English, as it were, is the only mainstream language with the absence of morphological markers for conveying T/V relationships; i.e. the verb is not marked for this purpose. In other words, a lexeme (i.e. pronouns corresponding to French *tous* and *vous*) and an inflectional morpheme do not reside in the inventory of the language to serve that purpose. This inevitably implies that in English the conceptualization of social deixis may as well be different. A native speaker of English, naturally, is aware of social deixis, yet expresses it non-consciously. Our reading is that the understanding of social distance is embedded in the users' mind-set, yet it is non-conscious; non-conscious in the sense of "…'unspoken rules' when it comes to social interaction … grammatical, semantic (or syntactic) rules, idiomatic

conventions, innumerable idiosyncrasies and specific linguistic customs that arise from one's local environment" that "...we must know in order to communicate effectively with those around us" (Williams, online).

Non-conscious usage can be approached through the notion of speaker meaning (Haugh 2013). It can be understood not only as a cognitive notion but also as a deontological notion: what the speaker is committed to in interaction (Haugh 2013: 41): "A speaker means something by intending that the hearer recognise what is meant as intended by the speaker," but also, the speaker is held accountable to the moral order for what he or she is taken to mean in interaction". The communicative intention encompasses the speaker's meaning and the hearer's interpretation of this meaning, for "... the speaker's stance and the hearer's role in the disambiguation of the message are crucial in making the meaning of the message" (Firbas In: Urbanová 2003: 14). Terminologically speaking, the meaning of the message can possibly be equated with the interactive meaning (on the term 'interactive meaning', *cf* Urbanová 2003). This is generated by the speaker-hearer interaction that indisputably discloses the speaker's attitude to the message, in other words, imparts the modal meaning of the utterance. A communicative intention represents "what the speaker is committed to, or taken to be committed to, in interaction" (Haugh 2013: 41). Since discursive commitment (on the term, *cf* Brandom 1994) becomes socially consequential, it abides by cultural norms, hence relies on conventionalized communication styles. For this reason, speaker meaning may as well embody social commitment, which as a matter of fact is culture-sensitive.

2. Methodology of research

2.1 *Motivation for research*

Social commitment is one of the troublesome areas for anybody involved in intercultural communication. A typical example is English and Slovak (together with other languages): the latter possess(es) pronominal and morphological means of social deixis; the former lacks such elements. Hence, the underlying question is: if a language lacks overt markers of social deixis (lexical or morphological), do native speakers of that language have such a frame in their minds? Other questions may well arise: what does 'commitment' imply with regard to social deixis and do English and Slovak commitment rules in expressing social deixis exist? This generates another question, namely: what are those conventional (or rather conventionalized, as it were) means that allow a hearer access the speaker's commitment? We are aware that the nature of this phenomenon requires that several language levels be taken into consideration (phonic, morphological, semantic, syntactic, and pragmatic levels); to start with, the present paper focuses on the phonic level.

The reasoning behind the research specifically on the phonic level with regard to the issue concerned is that some information in speech is independent of the words and their sounds, and this information is conveyed by intonation. Intonation represents at once discrete and categorial phenomena. The former (discrete components) are meaningless units; the meaning is mapped upon them by context in which they can function by themselves or in chunks. The latter (categorial or gradient components) convey meaning in that the speaker can select from several options; this aspect of intonation may well be related to physiological and psychological phenomena that Gussenhoven (2016) calls biological codes. Intonation is used to express a number of functions including attitudes – these appear to be directly linked to Gussenhoven's

biological codes (2016). Due to the absence of overt (explicit) markers of T/V distinction, it can be presupposed that English speakers rely on markers embedded in mind; logical inference is that markers of T/V distinction are present, yet in the form of non-conscious utilization of categorial aspects of intonation.

2.2 *Research problem*

As the two cultures under consideration use different ways of imparting interactive meaning of social distance/closeness deriving from the presence or absence of morphological and/or pronominal language means, a question arises how social commitment in L2 culture can be unveiled by a speaker raised in an L1 culture. A tentative solution might be looking for manifestation of certain patterns in the phonic make-up of the verbal production. Therefore, Hence, we attempt to explore the phonic realization of an utterance to find out whether or not prosody plays a role in rendering social deixis in English. Our premise is that prosody is non-consciously used to signal the "social polarity", thus, for native speakers, can be considered a covert means of expressing social deixis. The research task builds upon the idea that in English in order to substitute for the lack of overt morphological markers prosody is a useful variable. We are interested in what language means on the phonic level native speakers of American English employ to signal social deixis. Our aim is to find out if the correlation between social norms and phonic language means employed in the make-up of a speech event exists. A study has been conducted to observe if the sound realization of utterances used among speakers with T-relationship and among those with V-relationship is indicative of any tendencies. To our knowledge, the phenomenon of social deixis has been widely studied with regard to honorifics, pronominal means, address forms, or communicative strategies, however, no studies on prosody as a means of expressing social deixis are available

2.3 *Research Corpus*

Speech is a social product through which communication fulfills its functions. Modern analyses of language focus on the usage of a language defined by physical, social, cultural, etc. context and/or by the psyche of the language user. For the intended analysis, it is necessary to compile a corpus of language material reflecting the language dynamics and modality of interpersonal relations. The logical inference is that this is commonly present in audiovisual text. Anecdotal evidence is that readily accessible media dialogues that would be sources of "natural" language do not provide for a balanced share of T/V relationships. In other words, in television shows intended for entertaining (sitcom, drama, crime stories, etc.) the majority of relations are T relations, while V relations are in the in minority. On the contrary, for instance, political debates present V relations in majority, and T relations appear in significant minority. The solution to obtaining a source with a balanced share of relations would be online corpora. In that case, however, the data would not be contextually defined, hence robbed of natural environment, i.e. of the development of relations between speakers. For this reason, we opt for a film dialogue even though this necessitates extracting balanced subcorpora; typically the T subcorpus is to be accommodated to the size of the V subcorpus. The selected research material is an episode of the drama series *Gilmore Girls*. The language material used is a DVD recording of the episode Bon Voyage (2007, Gilmore Girls, Season VII, season finale), with running time 40 min 16 s.

Our aim is to observe the feasibility of interrelating prosody and attitudinal and social deixis in American English. The starting premise is that a phonic level has a share in signaling power and/or solidarity in English; hence, our focus is the correlation between phonic language means employed in the make-up of a speech event and a language user as a variable in a T or V communication event. To reveal and interpret the data we used phonetic measurements and sociolinguistic interpretation; the latter was based on coding that drew on ethnographic description. Phonetic study comprises the analysis of temporal, tone, and force modulation of speech signal (cf Sabol 2006), i.e. measuring intra-sentential pause, measuring pre- and post-pausal melody contours, and identifying pitch accent placement. Sociological factors are considered to provide for the sociolinguistic interpretation of the data and to interpret the share of the data obtained in communicating (in)formal relationship. The ethnographic description made available the information about the principles guiding the linguistic behavior of speakers of Slovak as their L1 (as a representative of languages overtly discriminating between T and V relationships) and that of speakers of English as their L1.

Information is transmitted from a speaker to a perceiver through a variety of means; supresegmentals play a major role in interpreting utterance meaning by virtue of their interrelation with extralinguistic reality. In their nature, they are attitudinally and emotionally loaded in particular communication situations (Kráľ 1984, Gussenhoven 2002, Gussenhoven 2015). Prosody-related phenomena do not result from articulatory activities; they arise from temporal, force, and tone modulation of the exhaled air stream (Sabol – Zimmermann 1984, Laver 1995). For the purpose of the present research, three elements were selected for further exploration: pause, tonic stress, and melody, i.e. the elements contributing to the unveiling of grammatical structure of an utterance. Pause (representative of temporal modulation) participates in the rhythmic patterning of speech and segmenting an utterance. Easy and smooth perception is conditioned by placing tonic stress (an element of force modulation) on the portion of a speech signal laden with most information, thus eliminating potential or real possibility of interpreting an utterance in a manner different from a speaker´s intention. The third suprasegmental fulfilling a grammatical function is melody (tone modulation) which differentiates utterances in terms of content and communication forces. Namely, we are interested in the melody of pre-pausal syllables since the melody of the final part of an utterance is the most sensitive and important portion (Sabol – Zimmermann 1984, Zellner 1994; Viola – Madureira 2008; Cruttenden 1997).

3. Research procedure and findings

3.1 Ethnographic description and sociolinguistic interpretation

English and Slovak exemplify languages respectively lacking or having morphological means of expressing social deixis; thus, social distance is viewed by native speakers of the two cultures distinctively. Assumingly, the lack of morphological means in English has prevented the present-day lingua-cultural behavior from detailed stratification of relationships. The pronominal and morphological expression of the stratification of the society is much simpler in English than in Slovak. For this reason, it is much easier to describe lingua-cultural behavior in English than in Slovak. The English pronoun 'you' referring to a single recipient has both

respectful and familiar readings. Based on anecdotal evidence, the verb associated with singular 'you' has neutral meaning; respectful and familiar meanings are mapped onto it by other language means within the immediate linguistic environment.

In the present-day Slovak language (or other languages sharing the T/V principle), the unmarked/default singular form adopted in encounters within family and among friends is the T-form; the unmarked singular form in encounters with strangers (regardless of age or gender) is the V-form. Other encounters when the T-form is used are child-to-child, teenager-to-teenager, or the elderly-to-children/teenagers conversations. The generally held opinion is that when a child is below 18, an adult can use 'familiar you', while a young person is supposed to adhere to 'respectful you'. Many a time, breaching the age limit of 18 years in the usage of T-forms is a sign of arrogance. The appropriate form depends on the level of acquaintance, level of familiarity, or gender (e.g. at work, it is more typical of men to start first-name relationship sooner than of woman-to-woman or man-to-woman relationship). The V-form is the sign of social distance among people; the T-form signals closer relationship between people. A T-form tends to be used for addressing all family members. Among strangers, the transition from a V-form to a T-form is a sign of a closer friendship, the expression of rapport or of a more relaxed relationship; in the case of work associates, the same might be applicable, yet with the necessary streak of superiority when talking business.

The transcript was studied so that T and V utterances were identified and grouped to two subcorpora. The focal language material is native English, i.e. the morphological form cannot be relied upon. The only agent determining the type of relationship is the social context; i.e. sociological parameters like age, gender, social role, or relationship. To identify the type of relationship, we used a coding system. Firstly, each type of the partnership was assigned a number; e.g. a family member – 1, a contemporary – 2, an acquaintance – 3, a stranger – 4). The second step was to tag each code with T-relationship or V-relationship. Due to the non-presence of such a pattern in English, the Slovak mindset (described above as ethnographic feature) was used to attribute the T and V relationships to the four codes based on the contextualized information on sociological parameters, like those above the length of acquaintance. The target corpus contained encounters between men and women, between two age groups (teenager vs. adult), and among speakers with social roles of family members, close friends, neighbors, and strangers. The final codes were, e.g. 1T, 1V. We counted the utterances coded as 1V – 4V, and thus established the subcorpus of V utterances (all V utterances were considered for research). The third step was to compile the subcorpus of T utterances; the number of T utterances was adjusted to the number of V utterances so that the two subcorpora were parallel in size.

The studied episode comprised the total of approximately 1014 utterances[i] featuring twenty-four characters. We identified 67 "relationships" and/or verbal encounters. Out of 67 encounters identified, 63 were considered as T encounters, only four were considered as V encounters. Some of the identified encounters represented communicative situations when one person spoke to a crowd (altogether app 10 in the studied episode). In the studied corpus, the ratio of T : V encounters and/or utterances was 94 % : 6 %, which made it obvious that T encounters prevailed. Hence, for the sake of the research, it was necessary to balance the size of the corpora to be studied, i.e. to accommodate the T subcorpus to the size of the V subcorpus (in that the major characters were identical in the two subcorpora). In the present experimental case study, the V-relations subcorpus had 101 utterances; accordingly, we adjusted the size of the T-relations subcorpus. The two sets of 101 utterances underwent phonetic analysis and sociolinguistic interpretation.

3.2 *Phonetic measurements*

3.2.1 *Intrasentential pause*
The utterances of the two subcorpora were first subjected to the analysis of within-speaker intrasentential pauses. The procedure of the measuring of intrasentential pause was as follows. First, a mono sound track was extracted and modified to uncompressed 16-bit/48kHz (wav format) resolution. Then from the transcript, relevant portions of utterances were selected, and from sound waves, pause duration was measured (in ms). The material was processed by Praat 6.0.14. By means of an oscilogram, spectrogram, and perceptual checks (see Figure 1), pause intervals were detected and measured; subsequently, the obtained data were put into MS Excel 2010 so that they could be statistically evaluated by the Statistica 13.1 software.
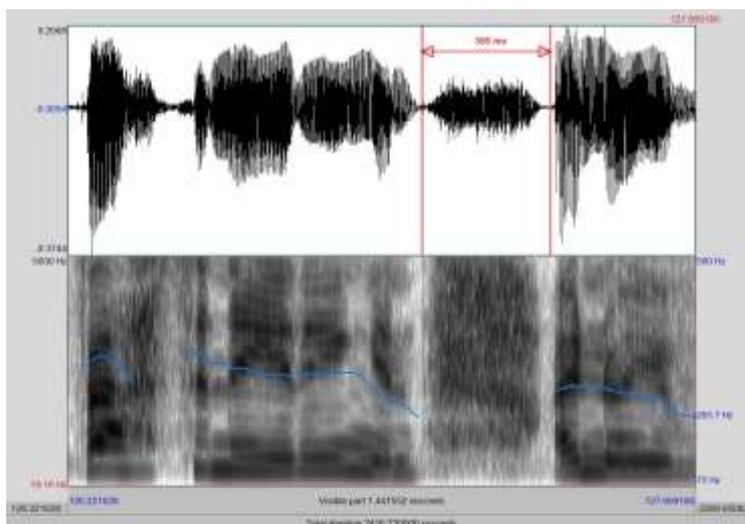
Figure 1: Oscilogram and wideband spectrogram in the Praat software with the melody
contour and the pause (duration: 305ms) in the phrase fascinating and//I know you´ve

In the V-relations subcorpus, we identified 70 intrasentential pauses; in the T-relations subcorpus, 44 intrasentential pauses. They covered both silent and filled pauses. Their duration and occurrence expressed in percentage is provided in Table 1 below.

Table 1: The distribution of intrasentential pauses in utterances with T and V forms

Utterances with T/V forms	Total of intrasentential pauses	Silent pauses		Filled pauses	
		occurrence	Median of duration	occurrence	Median of duration
T forms (N = 101)	44	93%	62 ms	7%	314 ms
V forms (N = 101)	70	91%	102 ms	9%	301 ms

The obtained pause measurements were statistically evaluated by the Statistica 13.1 software. By means of Shapiro-Wilk W-test for normality we found out that in both V-forms and T-forms silent pauses do not manifest normal distribution (p = 0), while filled pauses do (p = 0.17).

Figures 3 and 4 below present Box & Whisker Plots of the pause duration distribution in the investigated datasets generated by the Statistica 13.1 software.
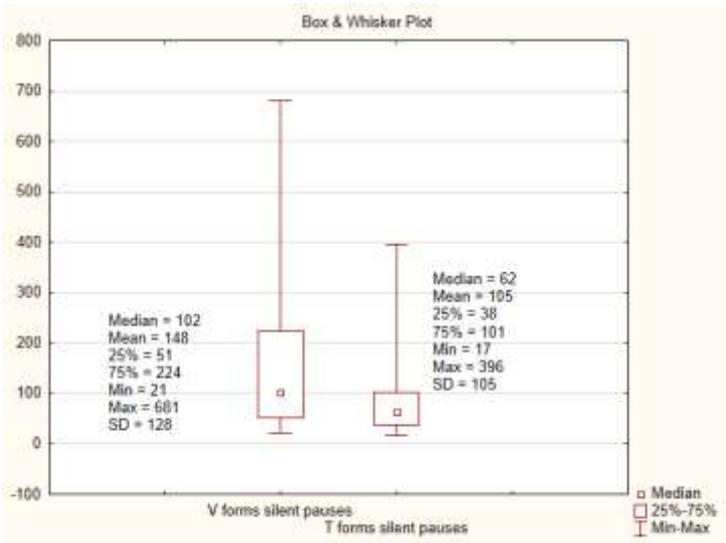
Figure 2: Box & Whisker Plot with median, mean, IQR, min, max and SD parameters of silent pauses in V forms and T forms (from program STATISTICA v. 13.1)
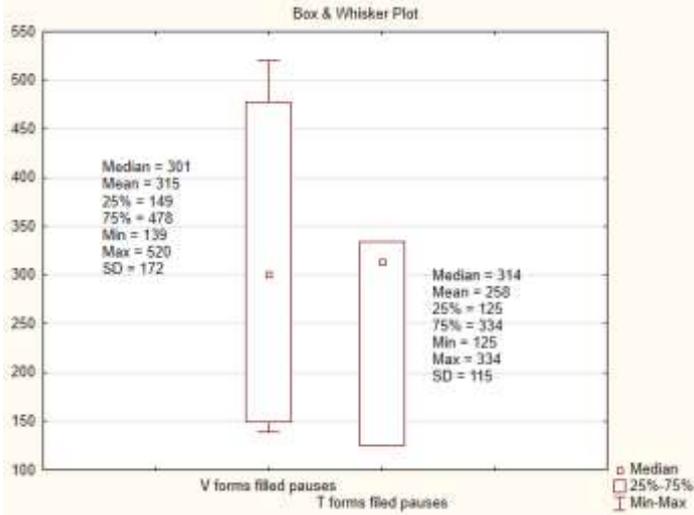
Figure 3: Box & Whisker Plot with median, mean, IQR, min, max and SD parameters of filled pauses in V forms and T forms (from program STATISTICA v. 13.1).

Based on the normality test outcome, in order to compare two independent datasets, we opted for non-parametric Mann-Whitney U test for two unmatched datasets; namely, silent and filled pauses. The statistical evaluation of the data ($p = 0.02$, $Z = 2.32$, $U = 959$ at the significance level of $\alpha = 0.05$) indicated that the difference in silent pause duration in utterances with T forms and V forms WAS STATISTICALLY SIGNIFICANT. The frequency of silent pauses is

presented in Figures 5 and 6 below. The pauses are divided into groups based on their duration. In the T-subcorpus, we identified 5 groups; in the V-subcorpus, we identified 7 groups.
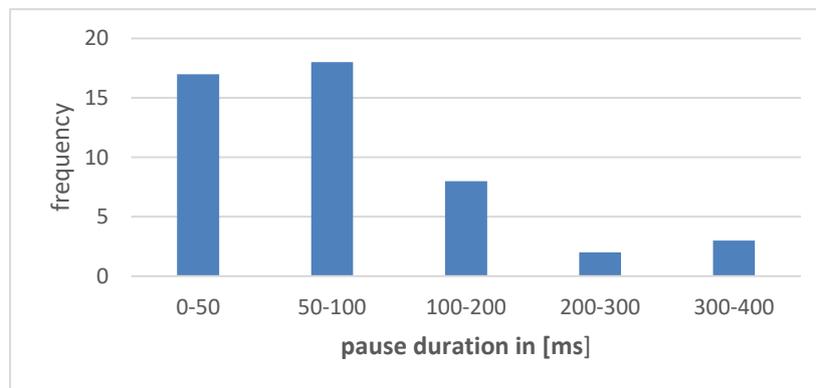
Figure 4: Frequency of silent pauses duration in T-forms
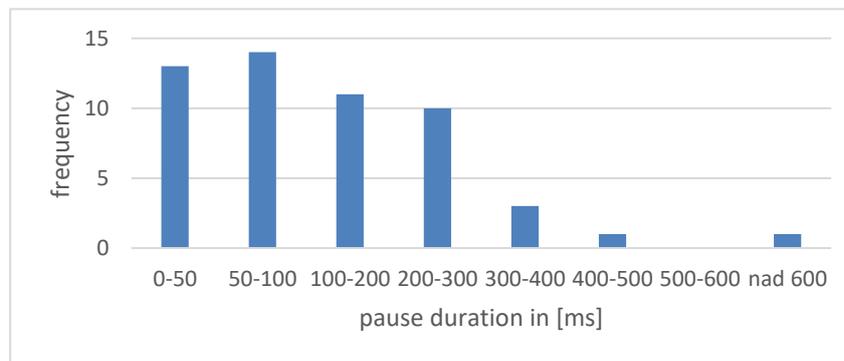
Figure 5: Frequency of silent pauses duration in V-forms

In the utterances with V forms, 9% of pauses were filled pauses (their duration was longer than that of silent pauses, the median of pause duration being 301 ms). In the utterances with T-forms, filled pauses represented 7% (their duration was longer than that of silent pauses, the median of pause duration being 314 ms). The median was used due to a low number of pieces of data (filled pauses in T forms: n = 3, filled pauses in V forms: n = 6). The difference in the duration of filled pauses in the T forms and of those in the V forms is not statistically significant at the 0.05 level of significance (Mann-Whitney U test: p = 0.52, Z = 0.65, U = 6). The results of the statistical evaluation may as well be influenced by the low number of the investigated data. For the same reason, in both T and V subcorpora, filled pauses fell into 1 group: in the former, that of 300-500 ms; in the latter, that of 100-600 ms.

3.2.2 *Pitch accent placement*

Pitch accent placement was detected through multiple perceptual checks by three independent assessors. The procedure involved the detection of points that, in an utterance, speakers emphasized due to their high informational importance, or implication of contrast and topicality. The classification included the following categories: broad focus, marked pitch accent placement, and the combination of the two. The corpus included utterances both marked

and not marked for special emphasis. The former included utterances with broad focus, utterances with marked pitch accent placement, and utterances with both broad focus and marked pitch accent placement in an utterance. Out of 101 utterances in the T-subcorpus and V-subcorpus, we identified the following results (Table 2).

Table 2: The ratio of utterances marked for special emphasis

TYPE OF SPECIAL EMPHASIS	T utterances	V utterances
broad focus (BF)	18	31
marked tonic placement (MTP)	10	3
BF + MTP in an utterance	14	13

The focal categories can be exemplified as follows. Small capitals denote pitch accent placement. Utterances 1 – 3 occurred in the T-subcorpus; utterances 4 – 6 in the V-subcorpus:

1. utterances with broad focus, e.g. LORELAI: WHAT are you TALKing about?
2. utterances with marked tonic placement, e.g. LORELAI: She's STAYing here?
3. utterances with the combination of BF + MTP in one utterance, e.g. MICHEL: I WANted to avoid YET aNOther emBARrassing INcident.
4. utterances with broad focus, e.g. LORELAI: HelLO, I'm LOrelai GILmore. I RUN the INN.
5. utterances with marked tonic placement, e.g. C. Amanpour: Do you know which ONE?
6. utterances with the combination of BF + MTP in one utterance, e.g. C. Amanpour: HI, ROry. How are YOU? NICE to MEET you.

The comparison of the two subcorpora revealed interesting findings. The striking difference is noticed in the occurrence of broad focus (18 utterances in the T-subcorpus vs. 31 utterances in the V-subcorpus) and in the occurrence of marked pitch accent placement (10 utterances in the T-subcorpus vs. 3 utterances in the V-subcorpus). A balanced occurrence is observed in the case of the combination of broad focus and marked pitch accent placement in one utterance (14 utterances in the T-subcorpus vs. 13 utterances in the V-subcorpus).

3.2.3 *Melody*

Pre- and post-pausal melody contours were analyzed as follows: melody contours were first obtained from Praat and then F0 values were measured in peaks of segmented syllables of words in relevant utterances (Figure 7a illustrates melody contours). The obtained data were subsequently put into MS Excel 2010 and statistically evaluated by the Statistica 13.1 software. Figure 7b below illustrates prosodic representation for the same utterance using a hierarchically structured set of phonological constituents (cf Gussenhoven 2015).
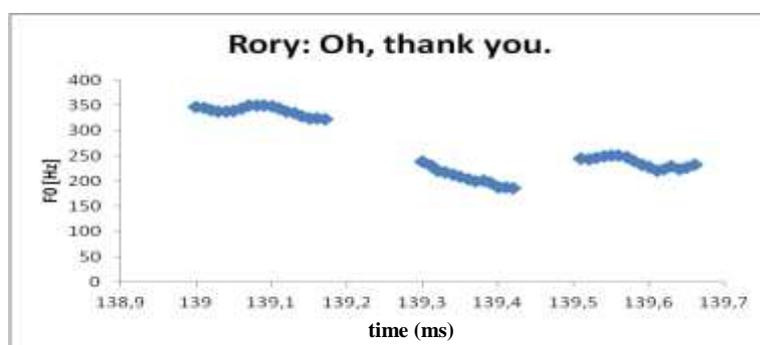
Figure 6a: Melody contour for the utterance: Oh, thank you.
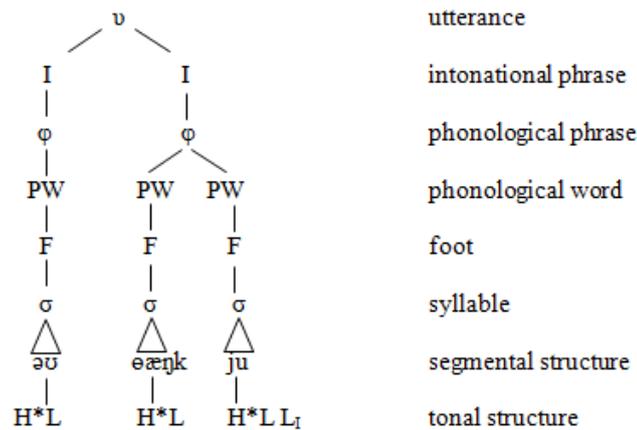
Figure 6b: Prosodic representation for the utterance: Oh, thank you.

We measured mean F0 values in semitones: these are more appropriate for investigating intonation since a human ear perceives sound frequencies logarithmically (Skarnitzl et al. 2016: 40). Mean F0 values for individual interactants in utterances with T-forms and V-forms indicate that according to non-parametric Mann-Whitney U test the differences are not statistically significant. However, some tendencies have been observed considering sociological parameters (see below).

3.3 *Discussion and implications*

3.3.1 Sociolinguistic interpretation of pause duration
The measurements imply that the interpretation of the obtained data necessitates considering the sociological parameter of AGE. The opposition of teenager (Rory) vs. adult (Lorelai and Luke) can be formed (Tables 3a, 3b). In the age category of 'teenager', shorter mean pause duration was observed in the V forms, longer mean pause duration in the T forms. By contrast, in the age category of 'adult', shorter mean pause duration was observed in the T forms, longer mean pause duration in the V forms.

Table 3a: Mean duration and median of intrasentential pause V and T utterances

Interactants	T-forms		V-forms	
	mean pause duration [ms]	median [ms]	mean pause duration [ms]	median [ms]
RORY (teenager)	196	196	172	134
LORELAI (adult)	122	110	139	89
LUKE (adult)	137	128	224	250

Table 3b: Mean pause duration and sociological parameter of age

AGE	mean pause duration in T encounters	mean pause duration in V encounters
Teenager (Rory)	longer	shorter
Adult (Lorelai, Luke)	shorter	longer

3.3.2 Sociolinguistic interpretation of pitch accent placement
In the utterances marked for special emphasis, the differences among the three characters are less significant, which is why sociopragmatic interpretation presented below focuses on the utterances not so marked. With regard to sociological parameters, also in this case the sociological variable of AGE seems to have interesting impact on the interpretation. Considering the three characters involved in both T and V relationships, the difference is between teenager vs. adult. In the case of teenager (Rory), the V utterances outnumbered the T utterances by 60%. In the case of adults (Lorelai and Luke), the situation was the opposite – more utterances not marked for emphasis occurred in the T relationships than in the V relationships; the difference in the ratio of T and V utterances can be attributed to the sociological parameter of gender only within one age group (adult): in Lorelai's case (female), the T utterances not marked for special emphasis outnumbered such V utterances by 80%, while in the case of Luke (male), the difference was only 42% (see Figure 8).
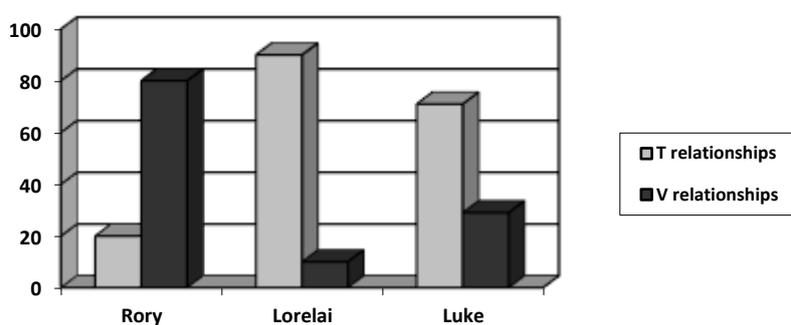
Figure 7: Utterances not marked for pitch accent placement and the sociological parameter of age

3.3.3 Sociolinguistic interpretation of melody
The comparison of melody contours in the dialogs forming the two subcorpora also necessitates taking into account the sociological variable of AGE. It is reflected in utterances by adult (Lorelai and Luke) vs. those by teenager (Rory), as illustrated by the tables below (Tables 4a, 4b). The age category of 'teenager' demonstrated lower mean F0 values in V forms, higher mean F0 values in T forms, while the age category of 'adult' showed lower mean F0 values in T forms, higher mean F0 values in V forms.

Table 4a: Mean F0 values in conversation employing T/V forms

AGE	T encounters		V encounters		V/T forms
	Mean F0 values [Hz]	Mean F0 values [ST]	Mean F0 values [Hz]	Mean F0 values [ST]	Mean F0 values [ST]
RORY (f)	277	17,63	249	15,79	-1,84
LORELAI (f)	246	15,58	253	16,07	0,49
LUKE (m)	162	8,35	191	11,20	2,85

90

Table 4b: Mean F0 values and sociological parameter of age

AGE	Mean F0 values in T encounters	Mean F0 values in V encounters
Teenager (Rory)	higher	lower
Adult (Lorelai, Luke)	lower	higher

4. Conclusion

The present paper focuses on the phonic level. The reasoning behind the research on specifically this level with regard to the issue of social deixis is that some information in speech is independent of the morpho-syntactic structuring and propositional content of an utterance – such information is conveyed by intonation. Intonation is used to express a number of functions including attitudes – these appear to be directly linked to what Gussenhoven (2015) calls biological codes. Due to the absence of explicit markers of T/V distinction, it can be presupposed that English speakers rely on markers embedded in mind; logical inference is that markers of T/V distinction are present, yet in the form of non-conscious utilization of categorial aspects of intonation.

With regard to the target language data, the interpretation needed to account for sociological parameters of AGE and GENDER. The phonetic analysis did not bring statistically significant results (except for one case), yet some tendencies were indicated. The only case of statistically significant results was the difference in silent pause duration in T/V utterances in that V utterances exhibited longer silent pause duration than T utterances. The frequency of pauses, intrasentential pause duration, melody contour, and pitch accent placement shared the tendency which can be explained through the sociological parameter of age. The age group of teenager showed the results reversed to those in the age group of adult: what was applicable to T and V in the teenager group, it was applicable to V and T, in that order, in the adult group. The sociological parameter of gender did not prove relevant.

Even though the conducted research is only done on a limited corpus of utterances and encounters, and further exploration on a larger sample of data is necessary to support the above findings, an important fact has been revealed. The type of encounters known as V in the cultures differentiating between T and V relationships is hard to identify in the Anglophone setting. This indicates that on one hand, a phonic analysis is one of relevant variables in the search for parallels in T/V and non-T/V cultures; on the other hand, an appropriate corpus for such a study needs a rich typology of encounters as T and V relations have proven to be transient in the Anglophone setting. For future corpus-based research, it is necessary to bear in mind that defining a relationship as T or V necessitates profound contextualization and abundance of speakers to assure a profile of a true T or V encounter.

It is an intricate task for a non-native language user of English coming from a T/V based culture to detect what is considered proper social manners and etiquette. We believe, this can be unveiled by research. The issues dealt with in the present paper highlight the areas worthy of exploration and show that social meaning appears to be embedded in language use and may be intricately interwoven with sociological variables that need further attention. This leads our thinking in the following direction: if phonic realization is taken as covert for a native speaker, yet possible to identify through phonic research, in the long run, its identification may be deemed potentially applicable for L2 acquisition, which may result as an overt language means in the non-native speaker's mind.

References

Boersma, Paul & Weenink, David. 1992 – 2016. *Praat: Doing Phonetics by Computer* [Computer program]. Verzia 6.0.14. Online program. 30 August 2016 <http://www.praat.org>

Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing and Discursive Commitment.* Harvard University Press.

Campione, Estelle & Véronis, Jean. 2002. 'A Large-Scale Multilingual Study of Silent Pause Duration' In: *Speech Prosody*, Aix-en-Provence, France. (http://www.isca-speech.org/archive) (Accessed 2016-05-18)

Cruttenden, Alan. 1997. *Intonation.* Cambridge: Cambridge University Press.

Crystal, David & Davy, Derek. 1973. *Investigating English Style*. Longman Group UK Limited.

Gussenhoven, Carlos. 2002. 'Intonation and Interpretations: Phonetics and Phonology.' In: *Speech Prosody. Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, ProSig and Universit ́e de Provence Laboratoire Parole et Langage. 47–57. (http://isle.illinois.edu/sprosig/sp2002/papers.htm) (Accessed 2016-09-15)

Gussenhoven, Carlos. 2015. 'Suprasegmentals.' In: James D. Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Vol 23. Oxford: Elsevier. 714 – 721.

Gussenhoven, Carlos. 2016. 'Foundations of Intonational Meaning: Anatomical and Physiological Factors.' In: *Topics in Cognitive Science* 8 (2016) 425–434, Wiley Periodicals, Inc. on behalf of Cognitive Science Society.

Haugh, Michael. 2013. 'Speaker Meaning and Accountability in Interaction.' In: *Journal of Pragmatics* 03/2013 48 (1). 41–56. (Accessed 2016-09-15) (http://www.sciencedirect.com/science/article/pii/S0378216612002925)

Kráľ, Ábel. 1984. *Pravidlá slovenskej výslovnosti*, Bratislava: SPN

Laver, John. 1995. *Principles of Phonetics (Cambridge Textbooks in Linguistics)*. Cambridge: Cambridge University Press.

Oravec, Ján & Bajzíková, Eugénia. 1986. *Súčasný spisovný slovenský jazyk: Syntax* [Present-day Standard Slovak Language: Syntax]. Bratislava: SPN.

Sabol, Ján. 2006. 'Pauza a dôraz.' In: Sabol, J., Bónová, I., Sokolová, M.: *Kultúra hovoreného prejavu.* Prešov: FF PU. 171 – 172.

Sabol, Ján & Zimmermann, Július. 1984. 'Komunikačná hodnota pauzy. ' In: *Úloha reči a hudby v životním prostředí. XXIII. akustická konference. 23rd Acoustic Conference on Physiological and Psychological Acoustics, Acoustics of Speech and Music.* České Budějovice: Československá vědeckotechnická společnost. 225 – 228.

Skarnitzl, Radek & Šturm, Pavel & Volín, Jan 2016. *Zvuková báze řečové komunikace. Fonetický a fonologický popis řeči.* Praha: UK v Praze Nakladatelství Karolinum.

Yule, George. 1986. *Pragmatics*. Cambridge: Cambridge University Press.

Urbanová, Ludmila. 2003. *On Expressing Meaning in English Conversation: Semantic Indeterminacy.* Brno: Masarykova univerzita, Brno.

Viola, Izabel C. & Madureira, Sandra 2008. 'The roles of pause in speech expression.' In: *Fourth International Conference on Speech Prosody.* Campinas, Brasil. 721 – 724. (aune.lpl.univaix.fr/~sprosig/sp2008/papers/id188.pdf) (Accessed 2011-02-05)

Williams, Robert 'Processing Information with Nonconscious Mind.' In: *Journal Psyche.* (http://journalpsyche.org/processing-information-with-nonconscious-mind) (Accessed 2017-02-22)

Zellner, Brigitte 1994. 'Pauses and the temporal structure of speech.' In: Eric Keller (ed.) *Fundamentals of speech synthesis and speech recognition.* Chichester: John Wiley. 41 – 62.

Language data compiled from Gilmore Girls, Season VII, DVD. Bon Voyage (season finale). 2007.

Assoc. Prof. Magdaléna Bilá, PhD.
Institute of British and American Studies
Faculty of Arts, University of Prešov in Prešov
17. novembra 1
08001 Prešov
magduska_bila@yahoo.com

Assoc. Prof. Alena Kačmárová, PhD.
Institute of British and American Studies
Faculty of Arts, University of Prešov in Prešov
17. novembra 1
08001 Prešov
alena.kacmarova@unipo.sk

Marianna Kraviarová, PhD.
Institute of Slovak and Mass media Studies
Faculty of Arts, University of Prešov in Prešov
17. novembra 1
08001 Prešov
marianna.kraviarova@unipo.sk

---

[i] In our understanding, an utterance embodies the modification of a sentence (syntactic construct) and the situationally and contextually determined attainment of a speaker's communicative intention delimited by speakers' swap (Crystal – Davy 1973, Yule 1996). A sentence signifies a detached, independent (predicative or non-predicative) structure, relatively non-compositional in meaning, intonationally delimited, and with grammar-consistent word ordering (Oravec – Bajzíková 1986).

Structural patterns of postmodifier in Nigerian English noun phrase

Mayowa Akinlotan, Vrije Universiteit Brussel

The present paper discusses the occurrence, structure, and complexity of the postmodifier in the Nigerian English noun phrase (NP) showing tendencies for structural simplification. It also compares its findings with patterns in British, Ghanaian, Singaporean, Honk Kong varieties. The paper shows how variables representing syntactic function, register, and weight shed light on specific contexts where we might or might not find (1) NP with or without a postmodifier (2) a clausal or phrasal postmodifier, and (3) a simple or a complex postmodifier. In addition, the paper shows that the extent of variation among different varieties of English is dependent on variables crucial to the construction choices being investigated. For instance, in (1), a postmodifier is realised while no postmodifier is realised in (2). (1) My car which I just bought last week has been stolen, and (2) My car has been stolen.The NP in (1) is structurally complex because it realises the clausal type of post modifier, 'which I just bought last week'. Meanwhile in (2), the NP (the car) lacks a postmodifier. In other words, the occurrence viz-a-viz non-occurrence of a postmodifier contributes to the overall structural and semantic complexity of the entire noun phrases, irrespective of the syntactic positions of the NP. Quantitative analyses of 8897 NPs indicate that in Nigerian NPs, a postmodifier is more unlikely to occur (61%) than not (39%). Further analyses show that prepositional phrase (57%), rather than clause (32%) or adjective (9%) or adverbials (2%), is the most preferred structural postmodifier type. It is also shown that realised postmodifiers are more likely to be structured in two-to-four words (51%) than four-words above. As for the predictive strength of variables studied, syntactic function is found to edge register in asserting influence and explaining different scenarios and contexts where we might or might not find a postmodifier, together with its structural type and weight. In other words, register, which is reputed as a significant indicator of structural variation (Biber, 2007; De Haan, 1993; Schilk and Schaub, 2016) is outweighed by syntactic function. The study further attests that significant structural simplification is largely present in the postmodifier structure of the Nigerian English noun phrase.

Keywords: *postmodifier, Nigerian English noun phrase, New Englishes, register, syntactic function and weight, structural simplification*

1. Introduction

Several new varieties of English have empirically attested to the tendency of structural simplification hypothesis. However only little empirical evidence supporting the hypothesis has been provided, and more, especially from the postmodifier structural constituent in Nigerian English noun phrase is required (Gorlach 1998). The present study therefore contributes empirical evidence from the postmodification constituent in the noun phrase structure in Nigerian English, showing how the structure and complexity of postmodifier, together with relevant predictors, reflect the tendency of structural simplification hypothesis. Given that the postmodification slot is a slot within the noun phrase structure which potentially could be complex as possible, then this syntactic element becomes a good syntactic unit with which the tendency for structural simplification in new varieties of English can be measured.

Together with variables representing syntactic function, register, and weight, three issues; (1) occurrence/non-occurrence (2) structural type and (3) structural weight, all of which insightfully relate to the structure of postmodifier are comprehensively examined, in the light of structural complexity/simplicity characterising outer and expanding Englishes. As can be seen in (1) and (2), the occurrence/non-occurrence of a postmodifier within a NP structure might highlight the presence of structural complexity or simplicity. In (1), a postmodifier is realised but it is not realised in (2), which makes (2) simpler and shorter to (1).

(1)   *My car which I just bought last week has been stolen*

(2)   *My car has been stolen*

Furthermore, the structural type of postmodifier (that is, choices among clause, prepositional phrase, adjective, and adverbial as postmodifiers) also shows the extent to which complexity/simplicity is present, and how the structural simplification hypothesis is indirectly shown. In (1) the postmodifier is a relative clause, rather than a prepositional, adjectival, or adverbial phrase. Thirdly, the structural weight of a postmodifier (i.e. the measurement of the words length) is also investigated, expatiating on the findings in the structural type. Furthermore, in (1) the length is nine (9) words' length, which could be longer or shorter. These three phenomena are discussed in relation to three relevant variables representing register, syntactic function, and weight. In other words, these variables, which have been shown to be influential in structural choices (Akinlotan 2018, Schilk & Schaub 2016, Brunner 2014, Biber et al. 1998, and see also Chapter 4) will provide us with specific contexts on (1) where we might or might not find a postmodifier, (2) where we might or might not find a simple or a complex postmodifier, and (3) where we might or might not find a short or a long (i.e. one, two, three, four, or longer) postmodifier. In addition to explicating the extent to which tendency of structural simplification is manifested by the structure of postmodifier in Nigerian variety, specific scenarios explicating the nature of structure of postmodifier in specific contexts in the variety will also emerge, allowing specific comparisons and hypotheses for similar inner circle Englishes.

Applying quantitative method on 8897 NPs extracted from the corpus material in the written section of the Nigerian component of the International Corpus of English (ICE), the present study will show that a postmodifier is more likely to be omitted from the Nigerian noun phrase structure (61%) than it is to be realised (39%). Also, it will be shown that realised postmodifiers are more likely to be constructed as prepositional phrase, rather than as a clause, an adjective, or an adverb. Furthermore, it will also be shown that these realized postmodifiers are usually two-to-four words length, such that complex postmodifier is rare. Indirectly, the structural patterns found will thus shed light on the structure of the postmodifier (slot) in Nigerian variety, as well as showing the extent to which the structural simplification hypothesis is present in the variety, together with specific contexts provided by the independent behavior of the selected variables. In addition to showing tendency of structural simplification, Barlage's (2014) assertion that structural node correlates with words' length in measuring complexity will also be tested out.

2. Postmodifier and variables

The structure and constituents of NP have been shown to be dependent on many variables such as (1) the kind of register that realises the discourse/text being investigated, and (2) the syntactic functions that NPs perform, and (3) the weight of the different constituents that make up the entire actual NP. Given that the postmodifier slot is such a syntactic position that allows for a complex or simple structure, then different variables such as those itemised above will be influential on what structural choices are made. In addition, influences of these variables have been established in previous literature, and a review of their effects is provided in the following section. In the subsequent section, the predictive strengths of these factors as found in the literature are discussed. Also, on the basis of their accounts in these previous studies, I thus propose expectations of how these variables will influence choices in the present study.

2.1 *Syntactic function*

Several works, for examples Meunier 2000, Schilk & Schaub 2016, Chapter 4, have shown that the syntactic function a noun phrase performs within a clause structure influences its internal structure. Gisborne (2003) and Hudson-Ettle & Nilsson (2002) provided evidence about the relationship between syntactic positions and structure of the constituents of noun phrase within a clause structure. More specifically, Hudson-Ettle et al. showed that premodifier complexity is influenced by the syntactic position occupied by the noun phrase that realises the premodifier. Furthermore, Schilk and Schaub (2016) and Chapter 4 showed that noun phrase at the subject position in a clause structure is structured simpler to noun phrase at other syntactic positions such as subject complement, preposition complement, direct object, etc. Following Gisborne (2003), Hudson-Ettle and Nilsson (2002), Schilk & Schaub (2016) and Chapter 4, it implies that a simplified noun phrase is one with fewer structural constituents.

In other words, one or two structural elements (e.g. postmodifier) will not be realized in such simplified NP. Given that postmodifier is potentially the heaviest structural constituent within the NP structure, then subject noun phrase (which, according to Chapter 4, is likely to be structured simpler), is likely to be structured without a postmodifier. Therefore, on the basis of previous findings in Chapter 4, and Schilk & Schaub (2016), it can be expected that occurrence of postmodifier will be influenced by the syntactic position of the NP realising the postmodifier. Following findings in Chapter 4, one can expect postmodifier to occur infrequently within NPs that are found at the subject position in a clause structure, while it is expected that there will be a high frequency of postmodifier in NPs found at other syntactic positions. Similarly, one can expect simple postmodifier to associate with subject NPs, while complex postmodifier associates with NPs at other syntactic positions. Following this expectation, we can then expect that clausal postmodifier will associate with non-subject NPs, while phrasal postmodifiers of any type (e.g. adjective, prepositional phrase) will associate with subject NPs. This expectation is informed by and derived from findings in Chapter 4-7.

2.2 *Register*

Register, including its characteristics text type, and genre, has been established as an important variable in syntactic variation analyses such as in genitive alternation (Akinlotan 2016b,

Rosenbach 2002), in dative alternation (Bresnan et al 2007), in particle placement (Gries 2003), in noun phrase structures in new Englishes (Schilk & Schaub 2016), in definite article usages (Wahid 2013), and in noun phrase complexity (Akinlotan & Housen 2017). Furthermore, Biber et al. (1999), Halliday (1988), and Varantola (1984) have also shown the significant influence of register on the presence and absence of internal elements such as postmodifier within a NP structure. More specifically, Biber et al. showed that academic text type, rather than non-academic text types such as literary works, is more likely to realise complex premodifiers. This means that non-academic texts are more likely to construct NPs without a premodifier. Meanwhile, when such non-academic texts do construct NP with a premodifier, such premodifier is likely to be a simple-structured type.

The aforementioned characteristics of register and structural choices can be extended to the postmodifier. Schilk & Schaub (2016) and Chapter 4 have shown that there is a relationship between register and likelihood of of occurrence of different syntactic constituents (e.g. determiner, premodifier, and/or postmodifier) in different varieties of English. This implies that the internal structure of NPs can be predicted on the basis of the texts that realise them. In Chapter 4, it is found that postmodifier rarely occurs in certain texts such as interactional, literary, and student essays. Relatedly, Jucker (1992), who studied the internal structure of NPs in relation to text types (upmarket versus down market newspapers), found a significant relationship between the structure of NPs and the type of text that realises them. Jucker showed that up-market newspapers realise NP structural patterns that differ from those that are realised in down market newspaper. Specifically, down market newspapers realised more noun and names in prehead position than found in upmarket newspapers.

In the light of Jucker's sub-categorisation of media language into a social stratum of up-market and low-market newspaper, one can further understand the extent to which register's influence can go, in terms of explaining variation underlying structural patterns. Following previous studies, one can then expect a significant relationship between register (or text type, or genre) and occurrence, structural type, and complexity of postmodifier in our variety. On the basis of previous findings, one can expect a higher occurrence of post modifier in academic texts than in other text types. In the same vein, it can be expected that complex postmodifier (i.e. of clausal type and of long length) will associate with media, academic, and popular texts. Furthermore, interactional, student, business, and literary texts are expected to associate with non-occurrence of postmodifier, and when occurred, should associate with simple-structured postmodifier. More specifically, Biber (1999)'s assertion that the occurrence of postmodifier (and premodifier) is about the same frequency in media texts will be tested out in our variety.

2.3 *Syntactic weight*

Syntactic weight has been found to be influential in different constructional choices and realizations. The overall structure of a noun phrase, which can be generally described as a simple or a complex NP, has been found to be influenced by the syntactic weight/length of structural constituents or elements that make up the entire NP structure (Hawkins 1994, Bresnan et al. 2007, Wasow 1992, 2002). For instance, Chapter 4 shows that the structural complexity of NPs in Nigerian English is influenced by the fact that certain structural constituents such as premodifier and/or postmodifier are likely to be omitted. However, when occurred, the structural complexity is simple-structured. Therefore, the omission of a

premodifier or a postmodifier constituent within a noun phrase structure will resultantly influence the weight of the actual NP that is realised. This means that an NP that consists of both a premodifier and a postmodifier at the same time is expected to be heavier than NPs that do not make use of both constituents or any of the either constituents. Therefore, the hypothesis of structural simplification in New Englishes can be explained in terms of syntactic weight of the overall NP, and also in terms of structural constituent's present.

This suggests that the structural weight of (other) constituents making up a noun phrase can explain the phenomenon surrounding the presence and/or absence of a postmodifier within an NP structure. Hypothetically, and following Schilk & Schuab (2016) and Chapter 4, the weight of a postmodifier, if there is any, is likely to be influenced by the weight of a premodifier, or a complement, or even a head noun. This expectation will be tested out in our corpus data. Following findings in Barlage (2014), Schilk & Schuab (2016), and Chapter 4, it can be expected that the weight in the prehead slot will influence the presence/absence of a postmodifier. Where a postmodifier is present, the weight in the premodifier slot is expected to associate with the weight of the postmodifier. Since each constituent within the NP structure, especially the premodifier and post modifier slots, has capacity for varying degree of syntactic weight, then it can be expected that some kind of relationship between the length/weight of premodifier and postmodifier in our corpus data will suffice. For instance, Barlage found that the weight of a postmodifier contributes much more to the complexity of the overall NP than the weight of a premodifier does.

2.4 *Occurrence/non-occurrence of postmodifier in varieties of English*

Data on the distribution of postmodifier in varieties of English is scanty. Until Schilk and Schaub (2016) put forward their detailed data and analyses, there was no work available showing the specific and predictive distributions and nature of the internal structures of noun phrase in new and established varieties of English simultaneously. The following distributions on the occurrence of the postmodifier in five different varieties of English are extracted from Schilk & Schaub's distributions which showed the presence/absence of internal elements in NP in four different text types representing academic humanities, conversation, social letters, unscripted speeches. The syntactic functions of the NP, subject versus non-subject, are also accounted for. Schilk and Schuab accounted for four levels of occurrence, relying on whether the actual NP consists of a premodifier and/or a posmodifier; (1) simple-NP, an NP which constructs neither a premodifier nor a postmodifier, (2) premodified-NP, an NP which constructs only a premodifier, but no other internal syntactic unit present (3) postmodified-NP, an NP which constructs only a postmodified, and (4) pre- and postmodified-NP, an NP in which both a premodifier and a postmodifier are present. In other words, (3) and (4) show presence/occurrence of a postmodifier, while (1) and (2) show absence/non-occurrence of a postmodifier. Following this interpretation, the following distributions on the occurrence and non-occurrence of postmodifier in five different varieties of English emerge. The five varieties of English represent Canadian (CAN), Hong Kong (HK), Indian (IND), Jamaican (JAM), and Singaporean (SIN).

Non-subject academic humanities

Occurrence	CAN	HK	IND	JAM	SIN
Postmodified	137	123	98	126	112
Unpostmodified	114	95	147	109	130

Subject academic humanities

Occurrence	CAN	HK	IND	JAM	SIN
Postmodified	35	46	47	58	52
Unpostmodified	114	136	108	107	106

Non-subject social letters

Occurrence	CAN	HK	IND	JAM	SIN
Postmodified	56	51	75	50	53
Unpostmodified	176	165	179	160	170

Subject social letters

Occurrence	CAN	HK	IND	JAM	SIN
Postmodified	14	9	13	13	15
Unpostmodified	154	175	133	177	162

3. Data selection and preliminary analyses

Noun phrases were extracted from the sixteen (16) different text categories in the written component in the Nigerian section of the International Corpus of English (ICE). In order to compare results to other varieties, NPs from media and academic texts are also extracted from Ghanaian and British varieties, using contemporary texts from BYU Corpus. Unlike ICE, BYU has an array of contemporary texts from a large number of varieties of English. A total of 1226 NPs from Ghanaian variety, 1072 NPs from British variety, and 3432 NPs from Nigerian variety is used in the final analyses. This leads to a grand total of 5730 NPs. Only academic and media texts in Ghanaian and British varieties are used. Table 1 shows the proportions of NPs that were extracted from each category, and how these various 16 texts categories in the ICE-Nigeria are reconceptualised into seven (7) registers. For example, the table shows that editorial and reportage text types make up the media register, while exams and student essays make up the student register.

The scores in the table show the size of NPs that each category realises. For examples, examination category realises 710 NPs while humanities in popular register realises 280 NPs.Each textual category consists of different text materials from which I selected the first set of texts. These first sets of texts are identified accordingly. The selected texts are stated as; AHum (1), ANsc (1), ASsc (1), ATec (1), Admin (1, 2, & 3), Business letter (1, 2, 3, & 4), Exams (1, 2, & 3), Novel (1 & 2), PopHum (1 & 2), Pop Natural Science (1), Pop Social Science (1 & 2), PTec (1 & 2), Reportage (1, 2, 3, 4, 5 & 6), Skills Hobbies (1, 2, 3, & 4), and Student essay (1, 2, &3). As can be seen it shows that only the first texts in Academic humanities, natural science, social science, and technical are used in the extraction process. Meanwhile, the first six (6) texts are used in reportage category.

Table 1: A description of text categories in the Corpus and the NPs extracted

GENRES IN THIS STUDY	TEXT TYPE IN NIGERIAN ICE		NPs
Student	Exams (686)	Student Essay (656)	1396
Media	Editorial (664)	Reportage (633)	1297
Academic	Humanities (308)	Natural Science (323)	1215
	Technical (279)	Social Science (305)	
Administrative	Business letter (694)	Administrative (609)	1303
Popular	Humanities (288)	Natural Science (267)	1206
	Technical (254)	Social science (397)	
Literary	Novel (1258)	---------	1258
Interactional	Social letter (609)	Skills hobbies (667)	1276

The extraction procedure followed revised procedure in Chapter 4 in which extracted NPs are those NPs that are syntactically interchangeable. That is, they can be substituted in their respective syntactic positions by a noun or a pronoun. NPs that are combined (Biber et al., 1999), such as *ministers and ambassadors* in *the recently appointed ministers and ambassadors of the Federal Republic of Nigeria* are identified as two constructions; (1) 'the recently appointed ministers of the Federal Republic of Nigeria' and (2) 'the recently appointed ambassadors of the Federal Republic of Nigeria'. Also, nominalised adjectives (Biber et al., 1999; Faragher et al. 2012), for examples, *the Nigerian, the sick, the rich, the masses, etc* are extracted. As Table 1 shows, a total of 9352 NPs emerged. These 9352 NPs are subsequently annotated for their structures and different variables. First, different NP realisations are identified. A revised version of Huddleston & Pullum (2002)'s NP theoretical framework is employed in order to identify and define a postmodifier. According to this framework, a postmodifier is a word or a group of word ranging from a phrase to a clause with a purpose of providing additional information about the head noun, such that the semantic identification of the head noun in the real world is obvious. In theoretical terms, this NP framework conceptually models NP as a functional category consisting of six (6) internal elements such as determiner (D) + premodifier (M) + head noun (H) + complement (C) + postmodifier (M) + peripheral dependent (PD).This framework contrasts with traditional descriptions of NP in that this framework does not only identify two structural nodes (i.e. Complement and Peripheral Dependent) but also theoretically distinguish between a complement and a head noun in one hand, and between a postmodifier and a peripheral dependent on the other hand.

In the present study, peripheral dependent (PD) is not identified as an independent syntactic element, but as a semantic realisation of a post modifier. In other words, our overarching NP framework (see chapter 2 for more argument on this reconceptualization) will suffice as: (D) (M) H C (M). In the present study, whether an NP is realized as D+H (the student), H+C+M (student of linguistics in year three), or as M+H+M (new President in old system) is noted and identified. This classification allows a clear picture of the use (occurrence and non-occurrence) of the postmodifier, such that a distribution of postmodified and unpostmodified NPs emerge. The syntactic function that an NP performs in the clause structure is also noted and identified. Eight syntactic functions, which follow from Chapter 4, are identified. Table 2 shows syntactic functions accounted for, together with illustrating

examples (see Akinlotan 2017, and Akinlotan & Housen for more on syntactic functions of the NPs).

Table 2: Syntactic positions and corresponding NPs within the clause structure

Syntactic Position	NPs within the clause structure
1. Subject	*The National Assembly* shall have power to make laws.
2. Subject complement	Drying of food crops is *an energy intensive operation.*
3. Apposition	The big three languages, *Hausa, Igbo and Yoruba,* dominate other minority languages in Nigeria
4. Direct object	Government parastatals lavish *huge sums of money* yearly
5. Indirect object	Experts are invited to *them* the many ways of doing business
6. Object complement	The Minister of Information called Boko Haram *a disease*
7. Preposition complement	Women with *mature skin* are as beautiful as the dewy youth
8. Adverbial	The room *downstairs* is being fumigated

The weight of the premodifier (preweight or prelength) and postmodifier (postweight or postlength) is also measured by counting the word(s) that make them up. Once counted, they are categorised as P (postmodifier) or M (premodifier), which can be P1 (one-word postmodifier; e.g *downstairs* in *the room downstairs*), P2 (two-worded; e.g. *running away* in *the man running away*), P3 (three words; e.g. *beautiful African attire* n *the lady in beautiful African attire*), P4 (four words; *who spoke fifteen Nigerian languages* in *the teacher who spoke fifteen languages*), and so on. The same counting method is applied to the premodifier, as in M1 (e.g. the *large* room downstairs), and M2 (The *beautiful African* dress).

In the counting method, prepositions, conjunctions, and determiners (except those that function as premodifiers (see Chapter 4, as well as Akinlotan & Housen 2017) were excluded. Structural type of post modifier (i.e. post type) are also identified as clause type (e.g. The man *who built the country* has died), adjective type (The money *available* is not enough), adverb type (The room *upstairs* is small), and prepositional phrase type (The problem *with Nigeria political system*). Finally, NPs are also categorised on the basis of the registers that realise them. The NPs are then categorised into the seven registers aforementioned. Having completed the annotation, the independent effects of each variable and the emergent patterns of structural distributions are analysed, using statistical methods of cross tabulation and chi square test, showing the relationship between constructional choices and variables representing syntactic function, register, and weight. In other words, only the results of the independent behavior of the variables are presented in the present study.

4. Results

The independent effects of the variables representing register, syntactic function, and preweight are presented here. How, and the extent to which the variables influence the use, structural types, and weight of postmodifier are shown. The distributions are followed by a chi square test of independence, showing what kind of relationship exists between the predictors and the issues at hand. The results are presented in this order: a general description of the structural distribution, which is followed by the independent effects of the variables.

4.1 *Overview of structural distributions*

Table 3: A table showing the overview distribution of NPs with and without postmodifier

NP with post modifier		NP without post modifier		Total (NPs)	
N	**%**	**n**	**%**	**n**	**%**
3432	**39**	**5465**	**61**	**8897**	**100**

As can be seen, NPs are more likely to be realised without a post modifier (61%) than with a postmodifier (39%). Given that post modifier can co-occur with different structural constituents within the NP, then it is important to know how the postmodifiers are distributed across different combinations. Hence, the table below shows the configuration of the post modifiers with different elements within the NP.

Table 4: A distribution of the co-occurrence of the postmodified NPs

NP structural types and examples	n	%
DHCM *The head of State in the last administration*	42	1
HCM *head of state in the last administration*	56	2
MHCM *Corrupt men of God in our society*	3	0
DHM *The head of the family*	1392	41
DMHCM *The sacked Secretary of State of the impeached government*	3	0
HM *Women in power*	809	24
DMHM *The graduate students of linguistics in year five*	701	20
MHM *Beautiful ladies in African attire*	426	10
TOTAL	**3432**	**100**

On the basis of our revised NP framework, only eight structural realisations are possible when a post modifier is present. All these eight possible realisations are realised/found in our corpus,

as can be seen in Table 4. However, there is a significant difference in the distribution of these structural realisations. As can be seen, a post modifier is likely to co-occur with determiner (40%), whereas the presence of a premodifier and a complement seem to impact on the realisation of the postmodifier. As can be seen, a postmodifier is unlikely to occur as a part of a very complex construction such as D +M+H+C+M. Given the above scenarios, it is then important to understand where we might or might not find a post modifier on the basis of the syntactic functions, preweight, and registers that characterise them. These are important variables that make our analysis more precise.

Table 5: Postmodifier occurrence and variables under study

| | Postmodified NP | | Unpostmodified NP | | Total | |
	n	%	n	%	n	%
REGISTER						
Academic	602	50	613	50	1215	100
Media	468	36	829	64	1297	100
Student	529	39	813	61	1342	100
Administrative	758	58	545	42	1303	100
Interactional	274	21	1002	79	1276	100
Literary	329	26	929	74	1258	100
Popular	472	39	734	61	1206	100
TOTAL	3432	39	5465	61	8897	100
SYNTACTIC FUNCTIONS						
Direct object	1074	91	109	9	1183	100
Indirect object	83	7	1138	93	1221	100
Object complement	13	4	357	96	370	100
Prepositional complement	851	97	28	3	879	100
Subject complement	479	33	962	67	1441	100
Apposition	54	2	2611	98	2665	100
Subject	795	78	222	22	1017	100
Adverbial	83	69	38	31	121	100
TOTAL	3432	39	5465	61	8897	100
PREWEIGT						
M0	2110	37	3574	63	5684	100
M1	1143	41	1652	59	2795	100
M2	153	43	203	57	356	100
M3+	26	41	36	58	62	100
TOTAL	3432	39	5465	61	8897	100

4.2 *Occurrence/non-occurrence of post modifier and its predictors*

In this unit more precision about the occurrence and non-occurrence of the post modifier is sought on the basis of the syntactic functions the NP perform, the weight of the premodifier (pre-weight), and the register (i.e. type of text) that realises these NPs. Rather than a general statement, the following distributions will reveal the underlying pattern characterising the use of postmodifier in Nigerian variety of English.

4.2.1 *Register and occurrence of post modifier*

Previous studies have shown register as a powerful variable explaining structural variation (Biber et al 1999, Schilk & Schaub 2016, see also Chapter 4). As can be seen in Table 5, the relationship between register and occurrence of postmodifier is weak $\{\chi^2 (6) = 516\ p<0.000\}$. However, the extent to which register explains the occurrence and/or non-occurrence of a post modifier within a noun phrase structure is still insightful. Media, student, interactional, literary, and popular registers are more likely to realise noun phrases without postmodifier (64%, 61%, 79%, 74%, and 61% respectively) than noun phrases with postmodifier (39%, 21%, 26%, and 39% respectively). On the other hand, administrative register is more likely to realise NPs with postmodifier (58%) than NPs without postmodifier (42%). Meanwhile the preference for a particular structural pattern does not emerge in academic register where the chance of producing NPs with postmodifier (50%) is just the same chance with producing NPs without postmodifier (50%).

In the distributions below, comparison between Nigerian variety and five other varieties is undertaken. As can be seen, there is a uniformity in pattern of occurrence/non-occurrence of postmodifier, such that occurrence/non-occurrence of postmodifier cannot be fully explained in terms of text type.

Social letter

	Postmodified		unpostmodified		Total	
	n	%	n	%	n	%
CAN	60	18	330	82	390	100
HK	60	15	339	85	399	100
IND	88	22	312	78	400	100
JAM	63	16	337	84	400	100
SIN	68	17	332	83	400	100
NIG[1]	**274**	**21**	**1002**	**79**	**1276**	**100**

Academic Humanities

	Postmodified		unpostmodified		Total	
	n	%	n	%	n	%
CAN	172	43	228	57	400	100
HK	169	42	231	58	400	100
IND	145	36	255	64	400	100
JAM	184	46	216	54	400	100

---

[1] Since social letter category is included in the Interactional register in Nigeria data, so we extract and compare distributions from this register. Also, this is the closet category in Nigeria data to social letter in Schilk et al.

SIN	164	41	236	59	400	100
NIG	**602**	**50**	**613**	**50**	**1215**	**100**

NPs without postmodifier are likely to occur in social letter, just as with academic text, though a varying degree of variation can be observed. In other words, social letters are more likely than academic text to realise NPs without postmodifier. While preferential pattern for absence of postmodifier in social letter is spread (e.g. 83% versus 17%), there is a small difference in academic text between the likelihood to realise NPs with or without postmodifier.

4.2.2 *Syntactic functions and occurrence of post modifier*

Chapter 4 and Schilk & Schaub (2016) have shown the important role syntactic functions play in explaining structural variation viz-a-viz occurrence and non-occurrence of a constituent within a noun phrase structure. As Table 5 shows, the relationship between syntactic functions and the occurrence of postmodifier is significant $\{\chi^2 (7) =5570\ p<0.000\}$; NPs at the indirect object, object complement, subject complement, and apposition positions are more likely to realise NPs without postmodifier (93%, 96%, 67%, and 98% respectively) than NPs with postmodifier (7%, 4%, 33%, and 2% respectively). Meanwhile, NPs at direct object, preposition complement, subject, and adverbial positions are more likely to realise NPs with postmodifier (91%, 97%, 78%, and 69% respectively) than NPs without postmodifier (9%, 3%, 22%, and 31% respectively).

Unlike register, syntactic function appears to be stronger in explaining different scenarios where we might or might not find postmodifier within a noun phrase structure. Schilk et al. (2016) shows how binary syntactic function can aptly present the influence of syntactic function on structural choices. Following this approach, a clearer picture of the strength of syntactic function emerges when a two-way dimension of subject versus non-subject syntactic positions is presented. Figure 1 shows a collapse of the eight syntactic functions into two syntactic functions; subject and non-subject NPs.
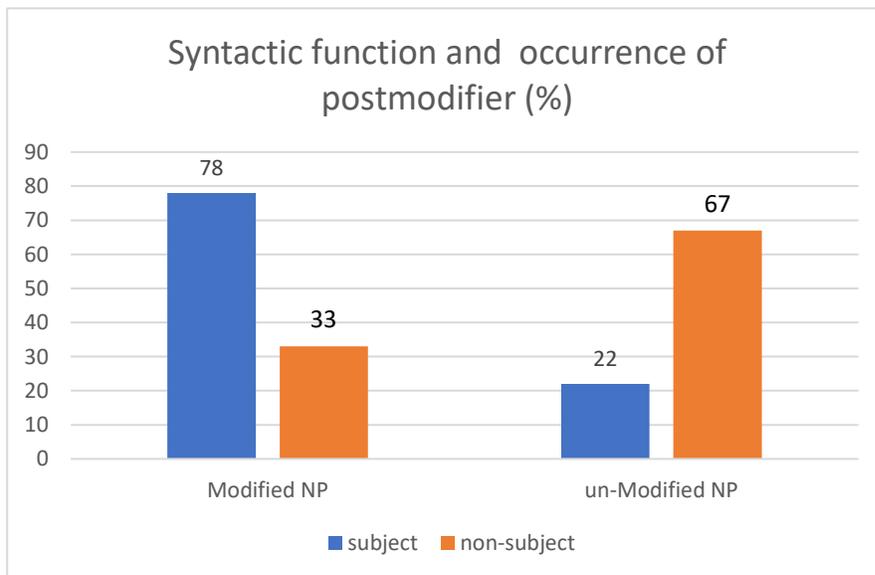
Figure 1: Relationship between binary syntactic function and postmodifier occurrence

In order to make a comparison between the influences of syntactic function on the occurrence of postmodifier in Nigerian variety and in other varieties studied in Schilk et al. (Canadian/CAN, Hong Kong/HK, Indian/IND, Jamaican/JAM, and Singaporean/ SIN), I restructure data in Schilk et al. and presented them below, alongside the distributions from Nigerian variety.

Subject NPs and occurrence of postmodifier in the varieties

	Postmodified		unpostmodified		Total	
	n	%	n	%	n	%
CAN	49	15	268	85	317	100
HK	55	15	311	85	366	100
IND	60	20	241	80	301	100
JAM	71	20	284	80	355	100
SIN	268	80	67	20	335	100
NIG	**795**	**78**	**222**	**22**	**1017**	**100**

Non-subject   NPs and occurrence of postmodifier in the varieties

	Postmodified		unpostmodified		Total	
	n	%	n	%	n	%
CAN	137	32	290	68	427	100
HK	174	40	260	60	434	100
IND	173	35	326	65	499	100
JAM	176	40	269	60	445	100
SIN	165	35	300	65	465	100
NIG	**2637**	**33**	**5243**	**67**	**1017**	**100**

As can be seen, syntactic function does not deeply explain the occurrence/non-occurrence of postmodifier in all of the varieties reported. Although there is some relationship, it is clear that non-occurrence of postmodifier is less sensitive to syntactic positions of the NP. Also, there are sharp similarities and dissimilarities among the varieties. Except for Canadian variety which is an inner circle variety, other varieties are outer circle varieties (Kachru 1985) and share some historical and developmental features (Schneider 2007), which expectedly should manifest similarities. However, Nigerian and Singaporean varieties are very much similar in that subject NPs are more likely to realise postmodifier, while the opposite is the case in Canadian, Indian, Hong Kong, and Jamaican varieties. In other words, subject NPs in Canadian, Indian, Hong Kong, and Jamaican varieties are more likely to realise NPs without postmodifier, while Nigerian and Singaporean varieties are less likely to omit postmodifier in their subject NPs. Furthermore, it can be observed that non-subject NPs and occurrence/non-occurrence of postmodifier appear to follow a uniform pattern in all of the varieties. Such is a scenario that further attests to a recurrent negative relationship between syntactic functions and occurrence/non-occurrence of postmodifier.

4.2.3 *Preweight and occurrence of post modifier*
Studies have shown that there is a relationship between syntactic weight and structural variation (Hawkins 1994, Wasow 2002, Akinlotan 2016b). Chapter 4 shows that the weight in the premodifier slot of a noun phrase contributes to the overall NP complexity. As Table 5

shows, the relationship between the weight of the premodifier (preweight) and the occurrence of the post modifier is weak {χ^2 (3) =14.62 p<0.002}; irrespective of the type of the preweight (M0, M1, M2, and M3+), noun phrases are more likely to be realised without postmodifier (63%, 59%, 57%, 58% and 58% respectively). When the noun phrase consists of one word of a premodifier (M1), the NP is very much likely to realise an NP without a postmodifier (59%). The M1 preweight preferential pattern is repeated for M0, M2, and M3 preweight which, respectively, return 63%, 57%, and 54% preferences for NPs without postmodifiers. The expected pattern is a strong relationship between M3+ preweight and NPs without postmodifiers. Also, expectation was for M0 preweight to have a strong relationship with unpostmodified NPs.

4.3. *Postmodifier structural type and its predictors*

Having shown where we might find or not find a postmodifier within the NP structure, then it is important to move closer to showing the structural types of the postmodifiers that are used. Postmodifiers can be realised as a clause, as a preposition phrase, an adverb, and an infinitive. On the basis of syntactic function, register, and preweight, insight into the structural type of postmodifier occurring with the NPs is provided. Given that clausal postmodifier is, potentially, the most complex postmodifier structural type, then a high frequency of this structural type will therefore attest to the nature of postmodifier complexity in our corpus. Since the data from the Nigerian variety will be compared with distributions from Ghanaian and British varieties, then results from Ghanaian and British varieties will be presented first.

4.3.1 *Postmodifier structural types in Ghanaian and British Englishes*

Table 6: showing distribution in Ghanaian variety by syntactic function and register

	Clause		Phrase		Adjective		Adverb		Total	
	n	%	n	%	n	%	n	%	n	%
Subject	141	27	324	62	51	10	3	0	519	100
Non-subject	144	20	541	77	19	3	3	0	707	100
Total	285	23	865	71	70	6	6	0	1226	100

	Clause		Phrase		Adjective		Adverb		Total	
	n	%	n	%	n	%	n	%	n	%
Academic	121	18	507	76	34	5	3	0	665	100
Media	164	29	358	64	36	6	3	1	561	100
Total	285	23	865	71	70	6	6	0	1226	100

As can be seen, postmodifier is more likely to be realised as a prepositional phrase (71%) than as a clause (23%), or as an adjective (6%). Postmodifier as an adverb (or adverbial phrase) is very much unlikely to occur, as the data shows. Furthermore, clausal postmodifier is more likely to occur in media text than in academic text. This is a pattern similar to distribution of phrasal postmodifier. Meanwhile, while clausal postmodifier is more likely to occur in subject NPs, phrasal postmodifier is more likely to occur in non-subject NPs.

Table 7: showing distribution in British variety by syntactic function and register

	Clause		Phrase		Adjective		Adverb		Total	
	n	%	n	%	n	%	n	%	n	%
Subject	121	33	221	60	21	6	5	1	368	100
Non-subject	275	39	391	56	19	3	19	3	704	100
Total	396	37	612	57	40	4	24	2	1072	100

	Clause		Phrase		Adjective		Adverb		Total	
	n	%	n	%	n	%	n	%	n	%
Academic	144	23	456	74	4	1	12	2	616	100
Media	252	55	156	34	36	8	12	3	456	100
Total	396	37	612	57	40	4	24	2	1072	100

Although at varying degree, British variety behaves somewhat similar to Ghanaian in that phrasal postmodifier is more likely to be used (57%) instead of a clausal postmodifier (37%). However, clausal postmodifier is more likely to occur in British variety (37%) than in Ghanaian variety (23%), while phrasal postmodifier is more likely to occur in Ghanaian (71%) than in British variety (57%). Moreover, while phrasal postmodifier (64%) is more likely than clausal postmodifier (29%) to occur in Ghanaian media text, the opposite is the case in British variety where clausal postmodifier (55%) is more likely than phrasal postmodifier (57%) to occur in media texts.In the following sections, results from Nigerian variety is presented, and compare to both Ghanaian and British varieties where obtainable. Table 8 shows distributions across variables representing register, weight, and syntactic function.

4.3.1 *Register and structural type of postmodifier*
The strong influence of register explaining variation in different languages and varieties have been established in the literature. As can be seen in Table 8, the relationship between register and structural type of postmodifier is significant {χ^2 (18) =257 p<0.000}. Interactional register, unlike academic (65%) and student 65%), is more likely to use a clausal postmodifier than any other structural type. It is noteworthy that academic (22%) is very much unlikely to realise clausal postmodifier. Given that clause is more complex than any other structural type possible as a postmodifier, then one would have expected such complex writing like academic to show more (or even the most) preference for clausal postmodifier.

On the other hand, academic and student texts (we had expected a significant variation between these two registers, given the different levels of proficiency exhibited in their writings, see Akinlotan 2016a) appear to be the registers most likely to use prepositional phrase postmodifier. Administrative text (62%) also follows academic and student texts in the choice for prepositional postmodifier. Furthermore, media (17%) and interactional (15%) texts emerged as the registers with the most likelihood to realise adjectival postmodifier, while student text is very much unlikely to realise the adjectival postmodifier type.

Table 8: Post-type and variables of register, syntactic function, and preweight

	Clause n	%	Prephrase n	%	Adjective n	%	Adverbial n	%	Total n	%
REGISTER										
Academic	133	22	389	65	46	8	34	6	602	100
Media	169	36	220	47	79	17	0	0	468	100
Student	167	32	346	65	16	3	0	0	529	100
Administrative	214	28	471	62	48	6	25	2	758	100
Interactional	142	52	90	33	42	15	0	0	274	100
Literary	126	38	171	52	21	6	11	3	329	100
Popular	157	33	261	55	45	10	9	2	472	100
TOTAL	1108	32	1948	57	29	79	79	2	3432	100
SYNTACTIC FUNCTIONS										
Direct object	324	30	658	61	64	6	28	3	1074	100
Indirect object	59	71	24	29	0	0	0	0	83	100
Object complement	8	61	5	38	0	0	0	0	13	100
Prepositional complement	283	33	474	56	68	8	26	3	851	100
Subject complement	198	41	239	50	29	6	13	3	479	100
Apposition	13	17	26	33	15	19	0	0	78	100
Subject	205	26	456	58	111	14	23	2	795	100
Adverbial	23	28	41	49	14	17	5	6	83	100
TOTAL	1113	32	1923	56	30	19	95	3	3432	100
PREWEIGT										
M0	661	31	1261	60	147	7	41	2	2110	100
M1	345	30	561	49	136	12	10	19	1143	100
M2	57	37	25	16	55	36	16	10	153	100
M3+	12	46	10	38	3	12	1	4	26	100
TOTAL	1075	31	1857	54	341	10	159	5	3432	100

More insight is provided by the use of adverbial postmodifier among the registers. As can be seen, media (0%), student (0%), and interactional (0%) texts show a knockout, which represents a complete negative relationship between adverbial postmodifier and these text types. Meanwhile, academic text (6%), unlike media (0%), student (0%), and interactional texts (0%), shows a positive relationship with the use of an adverbial postmodifier within a noun phrase structure. A number of similarities and dissimilarities emerge when distributions in Nigerian variety are compared with the distributions in Ghanaian and British varieties.

If we extract and sum up distributions from academic and media texts in Nigerian variety, the following scenarios emerge: clausal type (28%), phrasal type (57%), adjectival type (12%), and adverbial type (3%). This distribution thus implies that a relationship between register and postmodifier structural type in Nigerian variety is comparable to Ghanaian and British varieties, though in different respects. In terms of preferential pattern, Nigerian variety is more similar to Ghanaian variety (clausal 23%, phrasal 71%, adjectival 6%, and adverb 0%) than it is to British variety (37%, 57%, 4%, and 2% respectively). Such closer relationship between Nigerian and Ghanaian varieties is unexpected, as they belong to what Kachru (1985) described as outer circle. Also, Schneider's Dynamic Model (2007) aggregate them to belonging to the same phase. Although phrasal postmodifier is more likely than clausal postmodifier to be used in both Ghanaian and Nigerian varieties, some evidence of regional variation is still present, which is manifested in the preferential difference between 71% and 57% respectively.

On the other hand, while adverb postmodifier is very unlikely to occur in Ghanaian variety (0%), it is very likely to occur in Nigerian variety (3%), just as it is in British variety (3%). More specifically, academic texts across the three varieties behave similar. That is, phrasal postmodifier is more likely than clausal postmodifier to be used in academic text in Nigerian variety (65% versus 22%), in Ghanaian variety (76% versus 18%), and in British variety (74% versus 23%). As can be seen, there is very sparse variation in this respect. On the other hand, media texts across the varieties show a larger difference; for instance, clausal postmodifier is more likely than phrasal postmodifier to occur in media text in British variety (55% versus 34%), while the opposite is the case in both Ghanaian and Nigerian varieties. In Ghanaian and Nigerian varieties (i.e. outer circle varieties), media texts are more likely to use a phrasal postmodifier than a clausal postmodifier.

4.3.2 *Syntactic functions and structural type of postmodifier*
The syntactic position that a noun phrase occupies within a clause structure influences its structure and that of its constituents such as the premodifier and postmodifier (Schilk et al 2016). Also, stiff competition between syntactic function and register influencing constructional choices has been reported in Chapters 4-7. As Table 6 shows, there is a significant relationship between syntactic function and structural type of postmodifier {χ^2 (21) =173 p<0.000}; NPs at indirect object (71%) and object complement (61%) positions are very much likely to realise NPs with clausal postmodifier. Expectedly, subject NPs (26%), which are usually simple-structured, are very much unlikely to realise NP with clausal postmodifier.

Rather than clausal postmodifier, subject NPs seem to prefer a prepositional phrase postmodifier (58%) to any other structural type. The same pattern is exhibited in appositive NPs where a prepositional phrase postmodifier is the most preferred choice of structural type. Rather than a clausal postmodifier, NPs at direct object (61%), prepositional complement (56%), subject (58%), and adverb (49%) positions would also prefer a prepositional phrase postmodifier to any other structural type. While appositive NPs are very much unlikely to use clausal postmodifier, these NPs are the most likely NPs in all of the registers being studied to realise adjective type of postmodifier (19%). Next to appositive NPs is the adverbial NPs which stands at (17%) in the preference for adjective type of postmodifier. Next to adverbial NPs is subject NPs, which also show some level of choices in this direction. It is important to note that NPs at indirect object and object complement syntactic positions exhibit a negative relationship (0% versus 0%) with adjectival postmodifier.

110

In order to compare patterns in Nigerian variety to those of Ghanaian and British, distributions for syntactic functions other than subject function are collapsed and distributed as non-subject functions. In this way, equivalent scenarios emerge from Nigerian variety, allowing for a smooth comparison. The percentages of non-subject distributions in Nigerian variety which stand as clausal (34%), phrasal (56%), adjective (19%), and adverb (3%), shows that preferential pattern in Nigerian variety is more similar to British variety (39% clausal, 56% phrasal, 3% adjective, and 3% adverb) than to Ghanaian variety (clausal 20%, 77% phrasal, 3% adjective, and 0% adverb). On the other hand, postmodifier choices in subject NPs are very much the same across the three varieties (with clausal type of postmodifier standing at 33%, 27%, and 26% respectively for British, Ghanaian, and Nigerian varieties).

Furthermore, Nigerian and British varieties show some similarities in context where Ghanaian variety behaves differently. For instance, there is a clear-cut distribution in the use of adverb as postmodifier in Ghanaian variety, while British and Nigerian varieties show similar preferential patterns. In other words, while adverb is very unlikely to be used as a postmodifier in either subject or non-subject NPs in Ghanaian variety, it is likely to be used about the same chance in both Nigerian and British varieties.

4.3.3 *Preweight and structural type of post modifier*

Previous findings have shown that there is a relationship between syntactic weight and constructional choices (Hawkins 1994, Wasow 2002). As can be seen in Table 8, the relationship between preweight and structural type of postmodifier is significant $\{\chi^2\ (9) =278.5\ p< 0.000\}$; when there is no premodifier (M0), the postmodifier is very much likely to be realised as a prepositional phrase (60%). Meanwhile, prepositional phrase is very much unlikely to occur when there is premodifier of two-word length (M2). Where there is premodifier of two-word length (M2), postmodifier is more likely to be a clause (37%) or an adjective (36%) than it is to be a prepositional phrase. Some possibilities for complexity within the NP structure in our variety are shown by the fact that three or longer premodifier (M3+) is associated with the clausal postmodifier. In other words, M3+ is more likely to associate with clausal postmodifier (46%) than with prepositional phrase (38%) or with adjective (12%). Meanwhile, adjectival postmodifier is most likely to occur with an NP that has a two-word premodifier (M2). The same scenario is found with adverbial postmodifier, which is most likely to occur with two-word length premodifier (M2).

4.4 *Postweight and its predictors*

In this section, the relationship between post weight (i.e. the weight of the postmodifier) and the variables understudy (syntactic function, register and preweight) is presented. The distributions are presented in Table 9. Results from British and Ghanaian varieties are presented first, so that findings in Nigerian variety can be easily compared with them.

111

Table 9: Distribution of postweight by syntactic function and register in British variety

	P1		P2-P4		P5-P8		P9+		Total	
	n	%	n	%	n	%	n	%	n	%
Subject	98	26	199	54	73	20	1	0	371	100
Non-subject	198	28	399	57	102	15	2	0	701	100
Total	296	28	598	56	175	16	3	0	1072	100

	P1		P2-P4		P5-P8		P9+		Total	
	n	%	n	%	n	%	n	%	n	%
Academic	193	31	388	63	32	5	1	0	614	100
Media	103	22	210	46	143	31	2	0	458	100
Total	296	28	598	56	175	16	3	0	1072	100

As can be seen, complex postmodifier (0%) is very less likely to be used, while simple postmodifier (28%) is more likely to be used. If we further reclassify the distribution into a binary dimension, such that P1-P4 become simple and P5-P9 become complex, the pattern that emerged still shows that simple postmodifier (84%) is preferred to complex postmodifier (16%). If distributions between syntactic functions and register are compared to each other, then it can be observed that register explains the variation better than syntactic functions. For instance, it is clearly shown that P5-P8 postmodifier is more likely to occur in media than academic text. Whereas this is not the case with syntactic function where preferential difference is small. On the other hand, it is shown that longer postmodifier, P9+, is not related to register nor syntactic function. In the next table, the distributions from Ghanaian variety are presented and discussed accordingly, comparing scenarios where comparable.

Table 10: showing distribution of postweight by syntactic function and register in Ghanaian variety

	P1		P2-P4		P5-P8		P9+		Total	
	n	%	n	%	n	%	n	%	n	%
Subject	117	33	132	37	107	30	4	0	360	100
Non-subject	203	23	567	65	87	10	9	1	866	100
Total	320	26	669	55	194	16	13	1	1226	100

	P1		P2-P4		P5-P8		P9+		Total	
	n	%	n	%	n	%	n	%	n	%
Academic	191	26	445	61	92	13	5	1	733	100
Media	130	28	224	48	102	22	8	2	464	100
Total	320	26	669	55	194	16	13	1	1226	100

There is a similar pattern to British variety in that simple postmodifier is more likely to be used than complex postmodifier (P5-P9+). On the other hand, syntactic function in Ghanaian variety asserts more predictive influence on choices than syntactic function in British variety does. As can be seen, there is a relationship between syntactic functions and postmodifier

complexity; simple postmodifier of P2-P4 length is more likely to occur in non-subject NPs (65%) than in subject NPs (37%). Meanwhile subject NPs are more likely than non-subject NPs to realise simple-structured postmodifier of P9+ length. Furthermore, it is shown that media text is related to postmodifier of P2-P4 length (whereas P5-P8 postmodifier length is more likely to occur in media text than in academic text). Also, it can be seen that simple postmodifier is more likely to appear in academic text (61%) than in media (48%) text. In order to compare Nigerian variety to Ghanaian and British varieties, I present below the distributions from Nigerian variety.

Table 11: A distribution of postweight by syntactic function, register, and weight in Nigerian variety

	P1		P2-P4		P5-P8		P9 +		Total	
	n	%	n	%	n	%	n	%	n	%
REGISTER										
Academic	119	20	331	55	108	18	44	7	602	100
Media	128	27	301	64	29	6	10	2	468	100
Student	224	42	285	54	8	2	12	3	529	100
Administrative	285	38	286	38	93	12	94	12	758	100
Interactional	68	25	133	49	59	24	14	5	274	100
Literary	68	21	193	59	41	12	27	8	329	100
Popular	185	39	212	45	75	16	0	0	472	100
TOTAL	1077	31	1741	51	413	12	201	6	3432	100
SYNTACTIC FUNCTIONS										
Direct object	292	27	653	61	98	9	31	2	1074	100
Indirect object	39	47	35	42	0	0	9	11	83	100
Object complement	7	54	6	46	0	0	0	0	13	100
Prepositional complement	298	35	471	55	61	2	21	2	851	100
Subject complement	131	27	261	54	73	15	14	3	479	100
Apposition	14	26	31	57	6	11	3	6	54	100
Subject	271	34	461	58	45	6	18	2	798	100
Adverbial	36	43	47	56	0	0	0	0	83	100
TOTAL	1088	32	1954	57	29	49	96	3	3432	100
PREWEIGT										
M0	708	34	1184	56	175	8	43	3	2110	100
M1	358	31	568	50	138	12	79	7	1143	100
M2	33	22	96	63	11	7	13	8	153	100
M3+	9	35	17	65	0	0	0	0	26	100
TOTAL	1108	32	1865	54	324	9	135	4	3432	100

4.4.1 *Register and postweight*

Positive relationship (correlation) between register and structural complexity has been repeatedly reported in the literature. As can be seen in Table 11, the relationship between register and postweight in our corpus is significant $\{\chi^2 (18) =353 \ p<0.000\}$; literary, media and academia texts stand out in their preferences for two-to-four (P2-P4) postweight, which stand at 59%, 64%, and 55% respectively. Next to P2-P4 postweight in these three texts is the choice for a one-word (P1) postmodifier, which is the simplest structural type. Again, the variation in their distributions is minimal as preferences stand at 21%, 27%, and 20% respectively. Furthermore, the gap between literary (59%-21%), media (64%-27%), and academic text (55%-20%) is about the same. This shows that these texts exhibit similar pattern in their preferences for this sort of post weight.

Also, very noteworthy here is that literary text, which has been shown correlating with simpler or shorter NP (see Akinlotan & Housen 2017), is at par with academic and media texts, which, on the other hand, have been shown correlating with more complex NPs. Therefore, the expectation in which literary text is expected to behave differently from academic and media texts is not borne out in this case. Similarly, the texts which are most likely to realise one-word postmodifier (P1) are student (42%), popular (39%), and administrative (3%). It was expected that interactional text will behave similarly as literary text, attracting simple-structured form such as having a strong relationship with shorter length postmodifier (M1). Surprisingly, interactional text did not behave as expected by attracting simple-structured postmodifier. Instead interactional text turns out to be the text type with the most likelihood of producing more complex postmodifier (P5-P8). One would have expected academic and/or media text(s) to exhibit such association with (P5-P8) complexity. In a similar vein, administrative text (12%), rather than academic and/or media texs which are noted for structural complexity, is the text with the most likelihood for the most complex postmodifier weight (P9+). On the other hand, popular text shows a negative relationship (0%) with the most complex postmodifier.

The preferential pattern in Nigerian variety is similar to both Ghanaian and British varieties in the sense that simple postmodifier of P2-P4 is the most preferred structure, followed by P1 postmodifier, then P5-P8 length, with P9+ as the least used structure. Meanwhile, a closer look shows that Ghanaian and British varieties share a pattern in which Nigerian variety differs. While P2-P4 postmodifier structure is more likely to be used in the academic texts in both Ghanaian and British varieties, the reverse is the case in Nigerian variety, where simple postmodifier of P2-P4 structure is more likely to occur in media than in academic text. This sort of dissimilarity in pattern is also observed in P5-P8 complex postmodifier; while P5-P8 complex postmodifier is more likely to occur in the academic text in Nigerian variety, it rather is in the media texts in both Ghanaian and British varieties that we are likely to find such complex post modifier of P5-P8.

We can also note that academic text in British variety behaves slightly different from academic texts in both Ghanaian and Nigerian varieties in that P1 simple postmodifier is more likely to occur in British academic text than in media text; a scenario that is opposite in Nigerian and Ghanaian varieties, where P1 simple postmodifier is more likely to occur in the media than in academic text. However, it must be noted that this variation is spare in Ghanaian variety (26% versus 28%). Furthermore, complex postmodifier of P9+ structure is more related to Nigerian variety than it is related to British and Ghanaian varieties. While academic text (7%), rather than media (2%), in Nigerian variety, is likely to realise P9+ complex

postmodifier, the occurrence of such complex postmodifier in both Ghanaian and British varieties is almost non-existent, as no scenario is found in British variety.

4.4.2 *Syntactic function and postweight*
Syntactic functions have been shown providing insights into constructional choices. As Table 11 shows, there is a small but positive relationship between the syntactic positions of the NP and the weight of postmodifier within the noun phrase structure {χ^2 (21) =107 p<0.000}. Of all the different structures of postmodifier, the relatively complex postmodifier (P2-P5) is the most likely structure to be realised by NPs at the direct object position (61%), prepositional complement (55%), subject complement (54%), apposition (57%), while P1 postmodifier is the most preferred choice with NPs at the object complement (54%). Postmodifiers in NPs at the object complement position (54%) are more likely than NPs at other syntactic positions to realise P1 postmodifier. Furthermore, two-to-four words postmodifier (P2-P4) is significantly more associated with NPs at direct object position than with any other syntactic function. As can be seen, NPs at subject complement (15%) and apposition (11%) are syntactic positions with the most likelihood to realise postmodifier of five-to-eight words (P5-P8). NPs at the indirect object, object complement, and adverbial are very much unlikely to realise NPs with postmodifier constituting five-to-eight words. NPs at the indirect object position, which have a negative knockout relationship with five-to-eight length postmodifier (0%), is the syntactic function with which the most complex postmodifier is likely to be realised. As with five-to-eight words, NPs functioning as object complement, and as adverbial exhibit negative relationship with longer postmodifier (P8+).

As can be seen across the three varieties, the syntactic function of the NPs does not clearly explain the variation in the use of postmodifier as register clearly does. As the table shows, spare variation and clarity is observed in the relationship between syntactic position of the NPs and the distribution of postmodifier, which is the case in the three varieties. The differences in P2-P4 type of postmodifier in British variety (54% versus 57%), and Nigerian variety (58% versus 57%) are small. Although P2-P4 is the most used postmodifier structure across the three varieties, the pattern of variation in Ghanaian variety is more related to syntactic function than it is the case in Nigerian and British varieties. In other words, syntactic function does not clearly explain the distribution of simple P2-P4 postmodifier in British and Nigerian varieties, where there are respectively 3% and 1% difference between subject and non-subject NPs producing this type of postmodifier. Nevertheless, it can be observed that P2-P4 simple postmodifier is related to non-subject NPs in Ghanaian variety and is not clearly so in both Nigerian and British varieties which are very similar in the use of this type of postmodifier.

On the other hand, while subject NPs are more likely than non-subject NPs to realise complex postmodifier of P5-P8 in both British and Ghanaian varieties, the reverse is the case in Nigerian variety. Non-subject NPs are likely to realise such type of postmodifier, though with a small degree of preference. Similarly, except for Ghanaian variety, in both British and Nigerian varieties, simple postmodifier P1 structure is not clearly related to syntactic functions. Although in Ghanaian and Nigerian varieties, subject NPs are more likely than non-subject NPs to realise simple P1 postmodifier; a scenario that is the opposite in British variety, as the relationship is stronger in Ghanaian than in Nigerian variety. As found in variable representing register/text type, complex postmodifier of P9+ is more likely to be found in

Nigerian variety than in British and Ghanaian varieties, though this difference comes at a small degree.

4.4.3 *Preweight and postweight*

Although the influence of constituent weight has been shown on constructional choice, specific arguments about the relationship between weight in the premodification slot and weight in the postmodification slot have not been clearly put forward. As Table 7 clearly shows, there is a small but significant relationship between the preweight and postweight $\{\chi^2$ $(9) = 84$ p$<$0.000$\}$; two-to-four-word postmodifier is the overall preferred choice, while M3+ premodifier is the most preferred choice. One-word (P1) postmodifier also associates with different types of postmodifier at varying length. For instance, one-word postmodifier (P1) is more likely to be realised with three or longer length premodifier (M3+) than with two-word premodifier (22%).

The same pattern is found also with two-to-four postmodifier (P2-P4) which associates more with three-and-longer length (M3+) at 65% than with any other preweight (for instance, M0 stands at 56%, M1 at 50%, and M2 at 63%).However, and expectedly, there is a knockout (0%) in the relationship between M3+and P5-P8, and P9+, which shows that the more complex postmodifiers (P5-P8, and P9+) do not relate with three-and-longer preweight (M3+). Also, one-word preweight associates with the more complex P5-P8 postmodifier (12%) more than with any other premodifier structures (12% versus 8%, and 7%). Similar pattern emerges with the most complex postmodfier structure in which the most complex premodifier (M3+) has a negative relationship with the most complex postmodifier (0%). One would have expected the simplest and the shortest premodifier structure (M0) to associate with the most complex postmodifier (P9+), but this is not the case at 3%.

5. Interaction of predictors motivating postmodifier complexity

In this section the results of interaction of predictors motivating postmodifier complexity are presented. As can be seen in Figure 2, the interaction of register and syntactic function motivating simple or complex postmodifier is shown.

Postmodifier

Node 0
Category	%	n
■ simple	89,7	3077
■ complex	10,3	355
Total	100,0	3432

Register

academic; administrative; popular

Node 1
Category	%	n
■ simple	84,5	1547
■ complex	15,5	284
Total	53,4	1831

media

Node 2
Category	%	n
■ simple	93,9	681
■ complex	6,1	44
Total	21,1	725

literary; interactional; student

Node 3
Category	%	n
■ simple	96,9	849
■ complex	3,1	27
Total	25,5	876

Function

PrepComp; Subj; Appos; IO

Node 4
Category	%	n
■ simple	89,0	904
■ complex	11,0	112
Total	29,6	1016

DO; Adv

Node 5
Category	%	n
■ simple	83,0	478
■ complex	17,0	98
Total	16,8	576

SubjComp; ObjComp

Node 6
| Category	%	n
■ simple	69,0	165
■ complex	31,0	74
Total	7,0	239

Figure 2: A decision tree showing interaction of predictors motivating choices

As the decision tree suggests, there is some amount of interaction of the predictors in motivating simple or complex postmodifier. It can be observed that register explains more of the variation than syntactic function, which explains only 54% of the observations. More specifically, it can be observed that academic, administrative, and popular texts interact with prepositional complement, subject apposition, and indirect object syntactic functions.

6. Conclusion

In this paper, three issues relating to the structure of postmodifier in Nigerian variety of English in the light of similar varieties of English have been investigated. These include (1) the occurrence/non-occurrence of a postmodifier within an NP structure, (2) structural type of the used postmodifier within an NP structure being a clause, phrase, or word, and (3) the weight/complexity of the used postmodifier within the NP structure. Findings from Nigerian variety, in the light of other inner and outer circle varieties, show that Nigerian NP is more unlikely to realise a postmodifier within an NP structure (61%) than to realise one (39%). Given that postmodifier contributes to the overall complexity of NP, this result therefore shows that there is high tendency of structural simplification hypothesis being validated in the

variety. In other words, it indirectly attests to the hypothesis of structural simplification in New Englishes (Gorlach 1998, Schneider 2007, Schilk & Schaub 2016, Akinlotan 2016a, and also see Chapters 4-7). In addition to this, the recurrent preference for simple-structured postmodifier across the varieties further consolidates the tendency for simplified structural pattern in outer circle varieties of English. In most of the varieties studied, the use of complex postmodifier (P5, P9+) is rare, showing the nature of NP complexity in New Englishes, which Mesthrie et al. (2008) have also described as being likely to be simple-structured. As findings in Hong Kong, Indian, Singaporean, and Jamaican show, this manifestation of structural simplification in outer circle varieties are motivated by different factors, ranging from external and internal ones (Akinlotan 2017). As shown, patterns in Nigerian and Singaporean varieties behave similarly in the occurrence/non-occurrence of postmodifier. However, the Nigerian pattern differs from Indian, Jamaican, and Honk Kong varieties, which according to Schneider (2007), should exhibit similar features, given their similar stage of development. Such difference may demonstrate Babalola's assertion (2010) that Nigerian variety has tendency to manifest patterns of syntactic immaturity, which is expected to characterise a developing language.

Furthermore, the paper shows where we might or might not find postmodifier, on the basis of three determinants representing register, syntactic function, and preweight, which have been established in the literature as relevant to the phenomenon at hand. Biber et al (1998), Schilk & Schaub (2016), and many works have shown the significant predictive strength of register explaining structural variation. At the same time, Chapter 4, and Schilk & Schaub (2016), and Akinlotan (2017) have also shown that register and syntactic do compete for influences in motivating linguistic choices/alternation or variation. The present study further attests to the competition that associates with the predictive strengths of register and syntactic function. While register outweighs syntactic function explaining where we might find simple or complex postmodifier, syntactic function outweighs register explaining where postmodifier might occur or might not occur. This pattern is also found to be the case for the inner circle varieties (i.e. Nigerian and Ghanaian English) examined. Such result suggests that different contexts interplay with predictive strength of internal or external linguistic factors, together with the nature of linguistic variation being investigated (Biber et al 1998, Akinlotan 2017). This scenario reflects resultant issues that might be associated with a binary syntactic model methodology in which only two syntactic functions are distinguished (i.e. subject versus non-subject as employed in Schilk and Schaub 2016) In this model, non-subject function will consist of all other syntactic functions other than subject function, such that an imbalance distribution might occur.

Table 12 shows a detailed summary of the analyses, explicating how each predictor behaves in relation to occurrence/non-occurrence, structural types and structural weight of postmodifier. Among many other findings, Table 12 shows that administrative, literary and interactional registers, rather than academic and media, relate more with structural complexity than had expected on the basis of previous findings reported in the literature review section. As Table 7 and 8 show, administrative, literary, and interactional registers, rather than academic and media, are more likely to realise a varying degree of complex postmodifiers (P5-P8, P9+). Furthermore, the predictors exert varying degree of influence across the three issues investigated, such that there are patterns that emerge from each issue. For the postmodifier occurrence, we can see a pattern in which NPs in Nigerian English are likely to omit postmodifiers, which ultimately impact on the NP complexity. As for cases where we

might find a postmodifier, syntactic function appears the strongest predictor, followed by the register, and then weight.

For instance, the percentage differences[2] in syntactic function (82%, 86%, 92%, 94%, 34%, 96%, 56%, 38%) are not only higher than those in register (0%, 28%, 22%, 16%, 55%, 48%, and 22%) but are also equally divided along postmodifier and unpostmodified NPs, which allow us to make specific statements about specific contexts where postmodifier is present or absent within the NP structure. Meanwhile, there is stiff competition between register and syntactic function in asserting influence on postmodifier structural choices. As Table 12 shows, the percentage difference between syntactic function and register is very close. Yet, there is a pattern that emerges; prepositional phrase is the most likely postmodifier structural type, while adjective and adverbial structural types of postmodifier are always almost not preferred choices. As can be seen, it is noteworthy that complex premodifiers (M2, and M3) associate with clausal postmodifier. This suggests that, though in a very small way, there is still some sort of structural complexity present in the structural constructions of New Englishes, and indeed in Nigerian variety of English. Hence such presence of structural complexity, irrespective of how small it might be, may invalidate Babalola's (2010) assertion of characterising the syntactic structure of Nigerian variety as immature.

---

[2]These are numbers in bracket for each predictor. They are derived from deducting the highest distribution from the lowest/other scores. For instance, occurrence in academic returns 50% for postmodified and unpostmodified, which (50%-50%= 0%) returns 0%, while media (64%-36%) returns 28%. The next line of bracket contains percentage difference for structural type. For example, the percentage differences for academic, following the method above, returns 43% (65%-22%), 57%, (65%-8%), and 59% (65%-6%). The third line of bracket is for postweight percentage difference, which is derived from the same method. For example, student's percentage difference returns 12% (54%-42%), 52% (54%-2%), and 51% (54%-3%).

	Occurrence			Structural Type					Structural Weight			
	PM	**UPM**		**CL**	**PP**	**ADJ**	**ADV**		**P1**	**P2-P4**	**P5-P8**	**P9+**
REGISTER												
Academic (0) (43, 57, 59) (35, 37, 48)	+	+		-	+	-	-		-	+	-	-
Media (28) (11, 30, 47) (37, 58, 62)	-	+		-	+	-	-		-	+	-	-
Student (22) (33, 62, 65) (12, 52, 51)	-	+		-	+	-	-		-	+	-	-
Administrative (10) (34, 56, 60) (26, 26)	+	-		-	+	-	-		+	+	-	-
Interactional (58) (19, 37, 52) (24, 25, 44)	-	+		+	-	-	-		-	+	-	-
Literary (48) (14, 46, 49) (38, 47, 51)	-	+		-	+	-	-		-	+	-	-
Popular (22) (22, 45, 53) (6, 29, 45)	-	+		-	+	-	-		-	+	-	-
SYNTACTIC FUNCTION												
Direct object (82) (31, 55, 58) (34, 52, 59)	+	-		-	+	-	-		-	+	-	-
Indirect object (86) (42, 71, 71) (5, 47, 36)	-	+		+	-	-	-		+	-	-	-
Object complement (92) (23, 48, 53) (8, 54, 54)	-	+		+	-	-	-		+	-	-	-
Preposition complement (94) (23, 48, 53) (20, 53, 53)	+	-		-	+	-	-		-	+	-	-
Subject complement	-	+		-	+	-	-		-	+	-	-

(34) (9, 44, 47) (27, 39, 51)										
Apposition (96) (16, 14, 19) (31, 46, 51)	-	+		-	+	-	-	-	+	-	-
Subject (56) (32, 44, 56) (24, 52, 56)	+	-		-	+	-	-	-	+	-	-
Adverbial (38) (21, 32, 43) (13, 56, 56)	+	-		-	+	-	-	-	+	-	-
PREWEIGHT										
M0 (26) (29, 53, 58) (22, 48, 53)	-	+		-	+	-	-	-	+	-	-
M1 (2) (19, 37, 40) (19, 38, 43)	-	+		-	+	-	-	-	+	-	-
M2 (14) (21, 01, 27) (41, 56, 55)	-	+		+	-	-	-	-	+	-	-
M3+ (17) (8, 34, 42) (30, 65, 65)	-	+		+	-	-	-	-	+	-	-

Table 12: A summary of the performance of the predictors in the three issues discussed

References

Aarts, Fgam. 1971. On the distribution of noun-phrase types in English clause-structure. *Lingua* 26: 281-293. doi: 10.1016/0024-3841(71)90013-1.

Alo, Moses & Mesthrie, Rajend. 2008. Nigerian English: morphology and syntax. In Rajend Mesthrie (ed.), *Varieties of English. Africa, South and Southeast Asia*, 323-339. Berlin: Mouton de Gruyter.

Akinlotan, Mayowa. 2016a. The effects of structure and proficiency on determiner number (dis) agreement in Nigerian noun phrase. In: Antonio Moreno Ortiz and Chantal Perez-Hernandez (Eds.), *Proceedings of the CILC2016 8th International Conference on Corpus Linguistics*. Malaga: Spanish Association for Corpus Linguistics, 1-8. http://easychair.org/publications/paper/ (epic series in language and linguistics).

Akinlotan, Mayowa. 2016b. Genitive alternation in New Englishes: the case of Nigerian English. *Token: A Journal of English Linguistics* 5 (1): 59-73.

Akinlotan, Mayowa. 2017. Predicting Definite Article Usages in New Englishes: Variety Outweighs Genre and Syntactic Function in Nigerian English. *Anglica: an international Journal of English Studies,* 26 (2): 101-121.

Akinlotan, Mayowa & Housen, Alex. 2017. Noun phrase complexity in Nigerian English: Syntactic function and length outweigh genre in   predicting noun phrase complexity. *English Today,* 33 (3): 1-8. doi: 10.1017/S0266078416000626.

Anthony, Laurence. 2014. *AntCon (Version 3.4.3)* [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/.

Babalola, Emmanuel. T. 2010. *Nominal Group in Standard Nigerian and American English: A Systemic Linguistics Description*. Saarbrucken, Deutschland: Verlag Dr. Muller Aktiengesellschaft & Co. KG.

Behaghel, Otto. 1909. *Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern*. Indogermanische Forschungen 25: 119-142.

Brend, Ruth & Sullivan, William & Lommel, Arle (eds.). 2001. *LACUS Forum XXVIII: What constitutes evidence in linguistics?* Fullerton, CA: LACUS.

Berlage, Eva. 2014. *Noun Phrase Complexity in English. Cambridge*: Cambridge University Press. doi: 10.1017/cbo9781139057684.

Biber, Douglas & Johansson, Stig & Leech, Geoffrey & Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English* (9th ed.). Harlow: Longman.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: CUP.

Bresnan, Joan & Cueni, Anna & Nikitina, Tatiana & Baayen, Harald. 2007. Predicting the dative alternation. In: G. Bouma, I. Kramer & J. Zwarts (eds), *Cognitive foundations of interpretation*, 69-94. Amsterdam: Royal Netherlands Academy of Science

Brunner, Thomas. 2014. *Structural nativization, typology and complexity: Noun phrase structures in British, Kenyan and Singaporean English. English Language and Linguistics* 18, 23–48. doi: 10.1017/S1360674313000269

Brunner, Thomas. 2017. *Simplicity and Typological Effects in the Emergence of New Englishes. The Noun Phrase in Singaporean and Kenyan English*. Berlin/Boston: Walter de Gruyter.

Cayer, Roger & Sacks, Rence. 1979. *Oral and Written Discourse of Basic Writers: Similarities and Differences. Research in the Teaching of English* 13: 121-128.

Crystal, David. 2004. *Rediscover Grammar* (3rd ed.). Harlow, UK: Pearson Education.

De Haan, Pieter. 1993. Sentence length in running text. In Souter, C. & E. Atwell (eds.) *Corpus-based computational linguistics*, 147-161. Amsterdam: Rodopi.

Gisborne, Nikolas. 2003. *Relative clauses in Hong Kong English. World Englishes*19: 357–71. doi: 10.1111/1467-971X.00184

Gorlach, Manfred. 1998. Varieties of English world-wide: Where we stand. *Links & Letters* 5: 13-36.

Gries, Stefan.2003. *Multifactorial analysis in corpus linguistics: A study of particle placement.* New York: Continuum Press.

Hawkins, John. 1994. *A performance theory of order and constituency.* Cambridge: Cambridge University Press.

Hillier, H. 2004. *Analysing Real Texts: Research Studies in Modern English Language*. Basingstoke, England: Palgrave MacMillan.

Huddleston, Rodney & Pullum, Geoffrey. 2002. *The Cambridge Grammar of the English Language.* Cambridge*:* Cambridge University Press.

Hudson-Ettle, Dianna & Nilsson, Tore. 2002. Orality and noun phrase structure in registers of British and Kenya English*. ICAME Journal* 26: 33-62.

Jucker, Andreas. 1992. *Social Stylistics: Syntactic Variation in British Newspapers.* Berlin: Mouton de Gruyter.

Kachru, Braj. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In: Randolph Quirk & Widdowson, Henry (eds.) *English in the World*, 11-30. Cambridge: Cambridge University Press.

Kperogi, Farooq. 2015. *Glocal English: The changing face and forms of Nigerian English in a global world*. New York: Peter Lang.

Lamidi, Mufutau 2007. The noun phrase structure in Nigerian English. *Studia Anglica Posnaniesia* 42: 238-250

Mayowa Akinlotan
Department of Linguistics and Literary Studies
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussel
Belgium
mayowa.akinlotan@vub.be

Peter TRUDGILL

PS

Let us start with our traditional question. Why linguistics? What motivated you to study language and to deal with it professionally?

PT

From childhood I was always fascinated by linguistics even though, until I was twenty years old, I did not know there was any such thing. I read all the foreign language grammars in our local library, and collected as much information as I could find about the languages of the world. I studied French, German and Latin at secondary school, and asked for *Teach Yourself Spanish* and *Teach Yourself Malay* as birthday and Christmas presents. As – like other young people in the 1960s – I then hitch-hiked round Europe, I got interested in languages I encountered like Flemish, Luxembourgish, Plattdeutsch, Swiss German and Macedonian as well as Italian and Greek. When I left school I went to Cambridge University to study French and German, even though I was not a naturally gifted practical linguist and hopeless as a literary scholar. One of the most important things ever to happened to was therefore to find, quite by chance, in the
Modern Languages Library, a copy of Hockett's *A course in modern linguistics.* I realised, as I turned the pages, that this was what I wanted to study.

PS

You have been one of the most influential sociolinguists. Who influenced you at your beginnings and what determined the direction of your research?

PT

Without a doubt, William Labov. Again quite by chance, in 1967 I discovered in a bookshop in Edinburgh – where I was doing an MA in General Linguistics – the 1966 book *Sociolinguistics* edited by William Bright (the proceedings of the 1964 UCLA Sociolinguistics Conference). There I found Labov's article "Hypercorrection by the lower middle-class as a factor in linguistic change", which I thought looked very exciting. I bought the book, even though it was very expensive for a student, and took it home. I read Labov's article immediately, and decided straight away that this was the direction I would like my linguistic research to take.

PS

You are respected as, inter alia, an authority on dialects. Do you find any difference between research into 'languages' and 'dialects'? How did you select the dialects for your research.

PT

No, I do not find any difference here.
It seems mostly that dialects select me, rather than the other way round! For example, after spending some holidays on the south coast of Crete, I found myself becoming very interested in the local dialect and began to do research on that.

PS
What has been the role of field research in your research activities?

PT
Some of my work has not been based on field research but rather on library-based research, while other research has depended on field-work, for instance my early work in Norwich, later work on East Anglian English dialects, some work on Norwegian dialects, and work on the Cretan dialect just mentioned.

PS
In your view, is sociolinguistics paid relevant attention to in university curricula in Britain and other countries of your experience?

PT
It varies. In some places, the situation is very pleasing. In others, students seem to be remarkably ignorant about our subject. It could be better!

PS
Social and human sciences are rather underestimated in the present-day society. What do you think why? What do you think about the role linguistics should play in a society?

PT
Obviously linguistics should play a much bigger role in most societies than it does. There is so much ignorance about language amongst teachers, politicians and journalists – not to mention the general population – and so much unreasoning prejudice against nonstandard dialects, minority languages, and colloquial speech. Only linguistically well-informed thinking can counteract this.
The current predominance of neo-liberal business-school thinking, along the lines that everything is a business transaction and everyone is a "customer", does not help linguistics, nor the humanities in general, at all.

PS
Do you think that the advanced technologies have changed the linguistic behaviour of languages users? Do you consider this issue a big topic for sociolinguistics?

PT
Not very significantly - only in rather trivial ways. I do not consider it a big topic. It provides us with some good sources of data, maybe; but the uniformitarian hypothesis remains valid: technology has changed, but the nature of human languages has not; nor have the mechanisms of linguistic change. Moreover, the main means of transmission of phonological and grammatical changes remains face-to-face interaction.

PS
What are your main research objectives for the coming years?

PT

I am currently writing what I hope will be a sociolinguistic and historical-linguistic *magnum opus* on my native East Anglian English, a topic which I have devoted several decades of research to.

PS
And finally, could you please reveal something from your extra-academic activities? What are your hobbies? How do you relax?

PT
I write a weekly column on language and languages in Europe for the *New European* newspaper (this is relaxation!). I'm a great classical music fan. I support Norwich City Football Club. And I am very interested in cricket – I don't think Slovakia has a team? That's a pity.

Thank you very much for the interview.

<div align="right">Pavol Štekauer</div>