Towards a distribution-based corpus analysis of transfer-susceptible NP modifiers. A case of Polish advanced users of L2 English

Marcin Opacki
University of Warsaw, Poland
marcin.opacki@wn.uw.edu.pl

Romuald Gozdawa-Gołębiowski
University of Warsaw, Poland
r.gozdawa@uw.edu.pl

Abstract

The present study investigates selected transfer-susceptible constructions, comparing their distribution in the written output of Polish B2 and C1 learners of English against a corpus of examination essays from non-Polish users of L2 English and a parallel corpus of native English essays. Four functional morphemes have been chosen for analysis: the possessive *'s*, the genitive *of*, and two English articles, the indefinite *a* and the definite *the*. Our data consisted of a corpus of student writing ($N = 690$, of which $N = 150$ B2 were Polish learners of English, $N = 224$ C1 non-Polish learners of English, and $N = 192$ Polish C1 students, mean age = 20, $SD = 4$), supplemented by an NS reference group ($N = 32$). The analysis afforded significant variation among conditions, $F(6,2292) = 109.01$, $p < 0.05$. Post-hoc analysis showed that the mean score of the B2, L1-Polish group ($M = 5.58$, $SD = 4.11$) was significantly lower than that of the C1-Polish group ($M = 12.55$, $SD = 8.79$). However, no significant difference was observed between the C1 groups, with the C1-L1 non-Polish group ($M = 12.24$, $SD = 8.95$) having a similar score to its Polish counterpart. The frequency of all studied morphemes can be correlated with learner level, while any effect of L1 background remains negligible. This finding marginalizes the influence of transfer on the distribution of grammatical morphemes in

monitored L2 outputs and points to developmental sources[1] of morphosyntactic and lexical errors, perhaps along the lines of Collins at al. (2009).

Keywords: corpus linguistics; second language acquisition; second language teaching; second language learning; pedagogy

1. Introduction

Although transfer-based studies are a major area of SLA research and ever since Faerch and Kasper (1989) transfer has been recognized as an important communicative strategy in second language (L2) use, there is surprisingly little consensus when it comes to specific theoretical models, typologies or details of execution (cf. Master, 1987; Odlin, 2005; Otwinowska-Kasztelanic, 2016; for an overview and discussion). However, if we limit ourselves to the simplest scenario involving one native language and one foreign language, two assumptions seem fairly well established in the literature: the first language (L1) is the most probable source of transfer phenomena in any L2 (e.g., Ortega, 2009) and the likelihood of L1→L2 transfer is negatively correlated with the level of linguistic sophistication (e.g., more transfer at B2 than at C1), with the latter claim constituting the essence of the negative correlation hypothesis (cf. Bu, 2012; Bardovi-Harlig, 1998; Wishnoff, 2000).

While transfer remains a factor to be reckoned with in language learning, neither of the above assumptions is self-evident. On the one hand, evidence of persistent L1 influence in L2 written output is also documented, as argued by Wang (2001), Hussein & Mohammad (2011), or Jarvis and Pavlenko (2008), who argue for the increased presence of semantic transfer (as opposed to formal lexical transfer) at higher proficiency levels. On the other hand, transfer-free accounts of erroneous (non-target-like) L2 patterns have also been proposed and offer an alternative line of argumentation. For example, Collins, Trofimovich, White, Horst and Cardoso, (2009) in their insightful developmentally based analysis of "easy" versus "difficult" grammar constructions convincingly explain and operationalize morphosyntactic and lexicosemantic deviations from L1 norms in terms of perceptual salience, semantic scope and type frequency, with no reference to cross-linguistic influence (CLI).

Selecting between transfer-driven accounts and developmental accounts is much more than an academic challenge. The decision has far-reaching pedagogical consequences, as it entails radically different options for pedagogical intervention and should be based on solid empirical grounds. It is vital, therefore,

---

[1] We are grateful to an anonymous reviewer for insightful comments on transfer-free accounts of grammatical errors in L2 written production.

to develop some sort of a strict test to check for evidence of transfer, preferably corpus-based, as corpus tools and methodology are a perfect means to overcome some of the problems persistent in transfer-related research (e.g., distinguishing between language transfer and a lapse or simplification (cf. Jarvis & Odlin, 2000), or minimizing the impact of statistically insignificant findings in the data). In this paper, we propose such a procedure and illustrate its effectiveness using selected lexicogrammatical problems to identify potential transfer effects in the written output of Polish B2 and C1 users of English. As shall become transparent in the discussion below, the "L1→ L2 transfer check" is suitable even for small-scale pedagogical projects, providing language teachers with some rudimentary training in corpus construction with a unique tool to assess their learners' progress.

At the same time, we are not linking the concept of transfer to that of an error. We suggest that the analysis should be based on frequency counts of selected lexical items, without the researcher having to make *a priori* judgments about the grammaticality status of the forms under investigation (see below for details). This is an important advantage for pedagogically oriented studies, given the notorious problems involved in determining grammaticality (cf. Lewis 1993,1997; Widdowson 1994, 1997) and the different views on what should count as a norm in foreign language teaching (cf. the on-going debate about the status of native norms and native-speaker competence models, also see e.g., Jenkins, 2007; Holliday et al., 2015; Phillipson, 1992, 2001; Seidlhofer, 2006, and the discussion below).

This is intended to be a pilot study only, to be followed by an analysis of other problem areas and an overview of adequate pedagogical measures which would address the stipulated sources of errors and propose relevant remedial strategies. In the long run, our main objective is to show how corpora may be used to inform language teaching pedagogy (e.g., by providing learning-driven data to complement data-driven learning; cf. Beltz & Vyatkina, 2008) and to encourage more research in the area of corpus-assisted language teaching.[2] The paper looks at two areas of English grammar, which go beyond traditionally conceived morphology but which are readily perceived as susceptible to cross-linguistic influence, given the specificity of the Polish linguistic system – the genitive construction (*'s* vs. *of*) and the two articles (*the*, *a*). The genitive has been chosen because it represents what contrastive analysis dubbed "split", and assumed to be the height of L1→L2 interference: two items in the L2 corresponding to a single item in the L1 (Prator, 1967; Stockwell, Bowen, & Martin, 1965).

---

[2] We see innovative corpus-assisted language pedagogy as involving the use of target-language corpora by the learner to enable hypothesis testing and verification and the use of developmental corpora (in the sense of Belz & Vyatkina, 2008) for further practice and awareness building. Developmental corpora also give the teacher the input required for a focused reactive intervention (remedial action) , as opposed to "merely" proactive measures.

As for the article system, there is no lexical or grammatical equivalent in Polish to signal a comparable range of functional meanings (e.g., definiteness, countability, uniqueness, contextual relevance). This notable lack of the determiner category in Polish gives rise to an interesting academic dilemma; namely, whether or not absence of a feature is transferrable (cf. Arabski, 1979; Kellerman, 1987, for opposing views). This unclear status of article use in L2 English of Polish learners makes it an ideal testing ground for the purposes of transfer identification.

2. Error-based accounts of Interlanguage growth: Contrastive analysis and error analysis

By examining two aspects of nominal modification in English we are trying to put to the test some of the basic assumptions of the transfer-based analysis of L2 learning. As already mentioned, we propose to detect the presence of transfer without evaluating the grammaticality status of the forms under investigation. Instead, our analysis is based on frequency counts (i.e. distribution) of selected lexical items. To justify the importance of the shift from error-based to distribution-based L2 output analyses, let us briefly review the changing role of errors in applied linguistics since the downfall of the contrastive analysis hypothesis, henceforth CA (Lado, 1957; Ringbom, 2007).

CA failed because of its rigidly behaviorist claims, its inability to correctly predict the range of errors and non-errors in L2 production and to handle, or even predict, individual learner differences (cf. Schachter, 1974; Zobl, 1980; but see also James, 1980, for a more conciliatory view). The subsequent modifications of the system led to the development of error analysis (EA) (Corder, 1967), which shared one major feature with its predecessor – the perception of error both as a major indicator of language growth and a key factor in language assessment.

EA fell out of favor too, just as its predecessor did, and the reason for the growing dissatisfaction with error-related research seems to be ideological, as succinctly summarized in a series of talks by Steven Thorne and his colleagues (Thorne & Lantolf, 2004; Thorne & Payne, 2005, 2006): "contrastive analysis of actual use does not need to include incapacity evaluations of learners". "Incapacity" is a strong word and a judgmental one. Teachers who treat errors as manifestations of incapacity or students who believe themselves "incapacitated" because of a deviation from an L2 norm should clearly be advised to reconsider their attitude towards the teaching/learning process. There is no question of errors in L2 performance representing value judgments on the learner's personality or psychological integrity. In fact, one should remember that, in Corder's (1967) view, errors are both inevitable and necessary, and their presence may constitute evidence in favor of learning having taken place.

Another reason why error-based analyses are frowned upon is that even with a simplistic definition of an error as a "deviation from the norm" there needs to be a commonly recognized norm to serve as a reference point. With native standards repeatedly and increasingly questioned (cf. Jenkins, 2007; Mackenzie, 2014; Mollin, 2006), pedagogical norms are getting more and more elusive. Also, there is a growing interest in communicative effectiveness (e.g., degree of task completion) as an overriding criterion of successful language learning, with grammatical accuracy clearly relegated to the background.

Thirdly, the concept of a "language learner" has been replaced with that of a "language user" (cf. Cook, 2002), the difference being that *learners* remain error-prone, while *users* are allowed to experiment with language, based on their developing functional competence (cf. Kachru, 1985), thus enjoying the linguistic freedom otherwise associated with native usage. Teachers are encouraged to accept "errors" as "innovations or simplifications" in non-native Englishes, as can be seen in Seidlhofer's (2006, p. 47) *transdialectal enrichment, tolerance for diversity* in Jenkins (2007), or the *pluricentric approach* advocated in Jenkins (2006). Thus, depending on the currently predominant views on linguistic policies, transfer may be promoted or discouraged, in the hope of bringing the L2 production more in line with native speaker standards or – conversely – using it as evidence of emerging structural and morpholexical patterns of a semi-endonormative variety (in the sense of Kachru, 1985).

Finally, CA and EA excessively focus on errors and their typology, ignoring other factors (e.g., developmental ones) that trigger interlanguage development. This is why EA gave way to performance studies and interlanguage studies, which take a broader view of L2 competence. Importantly, engaging in contrastive L1 – L2 research is not tantamount to embracing CA; rather it reflects the belief that, for better or worse, cross-linguistic comparisons, perhaps in the form of Johansson's (1998a, 1998b) applied contrastive analysis are unavoidable in L2 pedagogy.

3. Introducing distribution-based analysis

A contrastive approach, based on the distribution of lexical items returns the information about the well-formedness or ill-formedness of a construction as a result of the analysis. In standard corpus-based research, error categories need to be specified before a corpus is created and all instances of erroneous sequences must be annotated accordingly, prior to running a corpus search. One of the most important advantages of a distribution-based corpus analysis is that it removes the unwelcome touch of subjectivity that most annotated learner corpora are tainted with, where outcomes depend to a considerable degree on *a priori* intuitive error mark-up decisions.

There are also fringe benefits of that approach – no painstaking mark-up of troublesome error taxonomies is required, which makes the whole process attractive for daily teaching-and-assessment purposes. Needless to say, it is imperative to retain the basic text structure annotation and the relevant metadata. If the distribution of a particular construction, phrase, lexical item, etc. in the corpus of Polish texts is to serve as a possible indicator of cross-linguistic influence having taken place, the results need to be juxtaposed against the data coming from non-Polish, L2 English users and a matching referential corpus of native English. Care must be taken to ensure maximum comparability across sub-corpora: as much as possible they should be matched for age, topic, purpose, level of formality, writing conditions, etc.

There are limitations to this approach, the most obvious one being the type of items that can undergo effective comparisons across sub-corpora. Following Mosteller and Wallace's early insights into stylometrics (Mosteller & Wallace, 1963), in this pilot study we limit the process of item selection to grammar words and grammatical (inflectional) morphemes rather than more meaningful content words. This is done mostly to counter the effects of topic selection on word distribution. As Mosteller and Wallace remark (1963, p. 275), "[t]he filler words of the language such as *an*, *of*, and *upon*, and, more generally, articles, prepositions, and conjunctions provide fairly stable rates (with which a word is used), whereas more meaningful words like *war*, *executive*, and *legislature* don't". Language teaching methodologies today will happily concur – after all it is a fairly well established contention (Benati, 2001, 2005; Lee & VanPatten, 2003; Lewis, 1993, 1997; VanPatten, 1996, 2003; VanPatten & Oikkenon,1996) that grammar words are indicative of more sophisticated, or more advanced, language use, as they reflect non-communicative system requirements, whereas learners consciously attend to meaning, rather than to form.

In this respect our study resembles frequency-based comparative morpheme studies, conducted in the past and methodologically founded upon frequency analysis (FA) (Cancino, Rosansky, & Schumann, 1978). Frequency analysis has shown itself to be quite a promising tool when it comes to establishing what learners do at particular stages of linguistic development through a rigorous process of measuring how the occurrence of specific grammatical items changes across a set period of time (Ellis & Barkhuizen, 2005, p. 93-95). Again, the advantage of this methodology is that is does not rely on EA and, therefore, does not necessitate an error taxonomy or typology. Rather than classifying items as errors and comparing their frequencies against the baseline of native speakers, we measure the frequencies of specific morphemes in native and non-native texts. In this way, our analysis becomes more data-driven. However, while what we propose is not longitudinal, but cross-sectional, we regard the number

of participants included in our project as one of its key advantages. Morpheme studies (e.g., Brown, 1973), while insightful, included relatively few participants.

Morpheme studies have been criticized at various points in history, most notably in the 1970s and 1980s (cf. Ellis & Barkhuizen, 2005, for an overview). Bley-Vromann (1983, p. 15) has accused the framework of falling victim to the *comparative fallacy*, in a manner not unlike error analysis. His primary issue with comparative, frequency-based research is that it relies on comparing non-native speakers with native speakers; and that this ultimately bears upon the development of interlanguage studies as an independent field of interest. While we acknowledge this point, we regard the possibility of making such a comparison as a strength of FA rather than a weakness. There is every reason to compare learner morpheme frequency longitudinally, on the assumption that learner language is a grammar[3] in its own right and of its own kind. However, we are of the opinion that comparing morpheme distribution between native and non-native populations, or any different types of populations for that matter, such as various types of non-native speakers, can be seen as benefitting interlanguage studies if only because such a comparison actually showcases how unique learner grammars are.

An undeniable advantage of FA, in line with Bley-Vromann's (1983, p. 15) point of view, is that it allows one to carry out interlanguage research based on phenomena that do not presuppose learner language inferiority or deficiency. Studies of learner language variation(e.g., Ellis, 1988) have shown that learner language differs from any given target language on counts other than merely a greater number of errors or impeded formal accuracy. Learners will use grammatical or "correct forms", but, due to an acquisitional (or learning) experience that is far removed from that of a native speaker, they will employ these forms in non-standard/non-native contexts. What is more, while both native and non-native speakers show considerable variation in language use, the native and non-native variation types are different (Ellis & Barkhuizen, 2005, p. 93-95). We remain convinced that all of these phenomena may be explored using research methods which incorporate some form of FA.

To sum up this section, frequency analysis has shown itself to be an effective tool for the study of second language acquisition. Not only has it contributed significantly to the discovery of universal patterns of L2 development, but it has also shown these patterns to be distinct from, and yet phenomenologically analogous to, L1 acquisition. This has been a major breakthrough in SLA studies (Lightbown,

---

[3] The statement is not the case of a sloppy metonymy. While it is certainly true that learner language HAS a grammar, it also remains true that it IS a grammar in the basic generative sense of rule-governed behaviour. We believe the assumption to be a necessary prerequisite for any study of interlanguage systems and their interaction with target and source languages, whether transfer-mediated or not.

1984; Lightbown & Spada, 1990) and it has shed some much needed light on the underlying mechanisms of learner language (Wode, Bahns, Bedey, & Frank, 1978, p. 184). The findings of FA are also of significance to second language pedagogy, since, in order to be effective, in-service teachers must factor in the unique pacing of the L2 acquisition process into their methods of instruction, taking into account all relevant transitional stages that build up to an adequate approximation of L1 knowledge that satisfies learner goals (Ellis & Barkhuizen, 2005, p. 109). The frequency-driven focus on filler words emerges therefore as a measure of linguistic competence, which we see as an exciting possibility, worthy of further research.

4. The study

4.1. Research objectives and rationale

4.1.1. The genitive

The genitive forms are a perfect choice for distribution-based research into transfer. In the first place, as already indicated, grammar words and grammatical morphemes yield themselves to a distribution-based analysis, unlike content words. Secondly, while CA itself is presumably beyond saving, some of its insights still linger around. Given the assumptions of the contrastive analysis hypothesis (cf. Lado, 1957; Prator, 1967), the genitive forms as used by Polish learners of English are potentially sensitive to transfer effects, because Polish makes no clear distinction between markers of the genitive and possessive cases. In both instances, the blanket genitive and possessive marker in Polish is inflectional and the endings signal more or less unambiguously the required syntactic relations and thematic roles. Thus *the destruction of the city* and *the city's destruction (by the enemy)* – to use a well-known Chomskyan example – will both be rendered into Polish by means of the same inflectional pattern– marking the word for *city* as genitive and putting it in the object (phrase – final) position – *zniszczenie miasta* (lit. *destruction city$_{GEN}$*). English, on the other hand, marks the possessive case with the Saxon genitive (the analytic genitive: X's Y) and the genitive case with the preposition *of*, yielding the periphrastic genitive (Y of X). Consequently, this typological asymmetry should result in Polish learners using the Saxon genitive to represent both the genitive and possessive cases ("narrowing" in the CAH terminology), accompanied by a concurrent underuse of the prepositional form. Underlying this prediction is the assumption that what actually transfers is the *morphological* mechanism of case-marking (cf. Jarvis, 2000).

On the other hand, transferring the Polish *syntactic* preferences should lead to favoring the periphrastic of-construction, as it matches the Polish word

order. Despite Polish being a relatively non-positional language, there is a strong preference among native speakers for the NP$_{GEN}$ to retain its post-head, phrase-final position. So, to return briefly to the example referred to in the preceding paragraph, *zniszczenie miasta* is the commonly accepted phrase, and the reverse sequence (*miasta zniszczenie*, lit. *city$_{GEN}$ destruction*) is heavily marked and restricted. Thus, our research will shed some light on the underlying tug of war between syntactic and morphological considerations in inducing transfer effects.

The periphrastic construction (X of Y) is also favored in the developmental (transfer-free) framework developed by Collins et al. (2009). The criterion of perceptual salience unequivocally points to *-of* as a more prominent item: an extra word in a written text, an extra syllable in pronunciation. As for semantic scope (i.e., the range of noun types that go with both patterns), the *-of* genitive is by far the less restricted one, as the only semantic category showing strong preference for the inflectional (*'s*) is the [HUMAN POSSESSOR]'s [OBJECT OF POSSESSION] – hence the contrast between *John's book* and *?*a book of John* (cf. Leech, 1994, for some inspiring discussion of the two genitive constructions). The criterion of type frequency (in native corpora) also favors the periphrastic *-of*: it shows up nearly four times as frequently as the inflectional (*'s*) in the British National Corpus (Davies, 2004).

To find out to what extent morphological and syntactic preferences have an impact on the distribution of the genitive forms and to establish evidence for or against transfer-related use of these forms we shall be addressing the following research questions:

a) Is the *-s/of* distribution pattern unique to Polish learners? Do learners from other linguistic backgrounds follow the same stages in the use of the two markers?

b) Does the pattern of *-s/of* distribution match the transfer-based hypothesis, that is, does the unbalanced (non-native-like) use of the two forms in the earlier stages of linguistic development lead to a more native-like use at later stages?

4.1.2. The definite and indefinite articles

Articles are prime candidates for transfer, both between languages that employ them and from an article-less language to a language with articles (cf. Ionin et al., 2008; Ekiert, 2004; Mayo & Pilar, 2008). The difficulties Polish learners face in the use of articles are attributed to interference in Śpiewak & Gołębiowska (2001, p. 174). Even though Arabski (1979) argued against recognizing article errors as transfer-induced in the case of Polish learners, as there was nothing to transfer, given the absence of the article category in Polish, Kellerman (1987)

counter-argued that the absence of a feature is just as transferrable as an overt feature, a view supported by Ellis (1994, p. 311-312).

In what follows we will use the distribution-based analysis to verify if the use of articles by Polish L2 users of English may be attributed to transfer. It is important to stress at this point that an error-based account of this particular grammar area would be considerably more subjective and time-consuming, as error categories overlap and the ultimate decisions very often depend on subtle contextual clues which may easily be misinterpreted by raters. Basing our predictions on the model developed by Collins et al. (2009),the definite article should have an advantage over the indefinite article, as it is more prominent phonetically and more frequent under normal discourse conditions (replacing the indefinite article, except for the "first-mention" contexts). For both articles, their overall frequency in English language corpora is staggering, and the model predicts no difficulty in mastering English articles by Polish learners. The frequency data are given in Table 1 below (all search instances were performed using recursive query syntax to exclude homographic, non-article uses of *a*, such as exclamations).

Table 1 Definite and indefinite article frequencies across various corpora with the same Tagset and annotation scheme

Article (corpus, tagset, citation)	Raw Frequency	% of entire corpus	Per 1000 words
a (BNC, CLAWS tagset, Davies 2004)	2054414	2.05	20.54
the (BNC, CLAW Stagset, Davies 2004)	5966774	5.97	59.67
a (COCA, CLAWS tagset, Davies 2008)	11663600	2.24	22.43
the (COCA, CLAWS tagset, Davies 2008)	28674422	5.51	55.14
a (GloWbE, CLAWS tagset, Davies 2013)	40025423	2.50	25.02
the (GloWbE, CLAWS tagset, Davies 2013)	106161644	6.64	66.35

A cursory inspection of the frequencies presented in the table reveals that the relative frequency of articles remains relatively invariant regardless of sample, which should come as no surprise given the widely accepted conjecture that word distributions in every language are Zipfian (Gelbukh & Sidorov, 2001; Zipf, 1935) in nature (i.e., that the frequency of each succeeding item is relatively half of the frequency of the previous item). One can state with a reliable measure of confidence that roughly twenty words out of a thousand are predicted to be indefinite articles, while as much as fifty words out of a thousand are predicted to be definite articles. This tallies with the predictions based on the criteria developed in Collins et al. (2009).

As for the level of difficulty of English articles for Polish L2 learners, the specific research questions take the form given below.

    a) Does the frequency of article usage match the transfer-based hypothesis, that is, is there a noticeable increase in frequency as the learner's competence develops, leading to a more balanced use at later stages?

b) Is there a difference between the frequency of use and patterns of change in definite and indefinite articles?

c) Is the pattern of article distribution across competence levels unique to Polish learners? Is the pattern of article distribution unique to Polish learners? Do learners from other linguistic backgrounds follow the same stages in the use of the two determiners?

4.2. Method

We accumulated a corpus of the written works of students (N = 690, mean age = 20, SD = 4) from different language backgrounds. The token count for the text samples in all sub-corpora ranged from T> 500 to T< 1000 (SD = 145.79) with GI (calculated as number of types over the square root of tokens multiplied by two, cf. Treffers-Daller, 2011) indices matched. We endeavored to control for content and format to the extent that this was possible; hence, we made efforts to match the papers we had available by topic and form. The forms that we selected for our corpus were: the opinion essay, the argumentative essay, the complaint e-mail, and the letter of complaint.

From among our data, 5 groups (N = 150) consisted of written text samples from B2-level students who sat the English Certification Exam at the University of Warsaw. These students were exclusively Polish, and all non-Polish works (i.e., those that came from foreign students studying in Poland) were removed from the sample to ensure representativeness. Seven of our groups (N = 224) consisted of written text samples from C1-level students who wrote practice application essays to various universities. These students were exclusively non-Polish and came from various language backgrounds, including French, Italian, Japanese, Chinese, and a number of others. The purpose of including these groups in the study was to establish which of the analyzed features are affected by transfer from Polish and which are not specific to the Polish context.

Six of our groups (N = 192) consisted of written text samples from C1-level students who wrote various papers as part of their Academic English writing class at the Institute of English Studies, University of Warsaw. These groups were included to inspect the influence of intensive language education on the frequency of the analyzed features. Finally, our last group (N = 32) consisted of native speakers of American English. This group also wrote university (college) application essays, which were structurally and thematically similar or identical to the papers from the previous groups.

4.3. Procedure

Our corpus was annotated for basic text structure and relevant metadata (student age, paper form, study level, etc.) using the 1.0 version of the Extensible Mark-Up Language (XML). We then conducted our analyses drawing upon a variety of software, including the commercially available *oXygen* XML editor as well as one concordancer (*AntConc* – Anthony, 2014a) and one word profiler (*AntProfiler* – Anthony, 2014b), using recursive query syntax whenever available.

For the first segment of our analysis, we compared the four relevant frequency features using a two-factor ANOVA assuming a standard α = 0.05, with group as the first factor and feature frequency as the second. Since the native speaker group was much smaller in comparison to the other groups, it was not included in this part of the analysis. Thus, at this point, we compared the essays of three randomized groups (N = 192, N = 192, N = 192) of learners, the B2Polish group, the C1 non-Polish group, and the C1 Polish group. Once statistically significant differences within learner groups have been established using Tukey's HSD test, we moved on to the second stage of our study.

In the second stage, we compared the structure frequencies between learner groups and the native speaker control group using a series of unpaired *t*-tests. Given that our control group was rather small and consisted of N = 32 samples, for the second portion of our analysis, we divided our data into groups of N = 32. We found that this division best suited the data that was available to us, forming groups that were representative and allowing for better data-cleaning (e.g., the removal of anomalous outliers such as incomplete essays, native specimens in non-native groups, etc.). In total, this division yielded 23 groups. We then performed *t*-tests for each relevant pair of data series.

4.4. Results

Using the analysis of variance described above, we compared the distribution of the four pertinent morphemes.[4] Groups represented language background (C1 with L1 Polish, C1 with L1 non-Polish, and B2 with L1 Polish) while treatments consisted of specific morphemes ([*a*], [*the*], [*'s*], [*of*]). Replications stood for individual students from among our population.

---

[4] Acting on a suggestion from dr. Agnieszka Leńko-Szymańska, we manually checked all the contexts involving the (*'s*) morpheme for irrelevant data (e.g., the syncretic *'s* marking the progressive aspect, as in *John's reading a book now*).
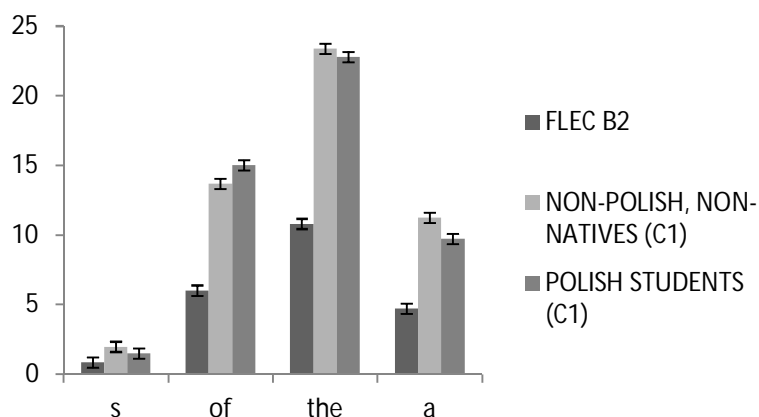
Figure 1 mean distribution of the analyzed morphemes across groups, with FLEC (Foreign Language Examination Corpus) data gathered from Polish students taking a certificate examination at the University of Warsaw[5]

In Figure 1 note the sharp discrepancy between proficiency levels, with Polish B2 learners of English avoiding the periphrastic genitive and the two articles, in marked contrast to the control C1 groups, both Polish and non-Polish. We take this tendency to be consistent with the negative correlation hypothesis (cf. the introduction to this paper and Bu, 2012), with a significant increase in frequency and more native-like use at later stages of linguistic advancement. At the same time, however, there is no evidence that the distribution of any of the four morphemes is unique to Polish learners of English, as both Polish and non-Polish C1 groups used them with comparable frequency, setting all C1 users apart from the B2 category. The data lends little support to the transfer-based account of the use of the genitive constructions by Polish learners, while remaining fully compatible with the developmental model: the periphrastic *of* remains markedly underused in comparison with the more advanced C1 group, as predicted by the negative correlation hypothesis but it is also more frequent in the B2 group than the inflectional (*'s*).

Since the analysis yielded significant variation among conditions, $F_{(6.2292)}$ = 109.01, $p < 0.05$, a further post-hoc analysis was warranted. For this purpose, we elected to employ the Tukey HSD test. The test indicated that the mean score of the B2, L1-Polish group ($M = 5.58$, $SD = 4.11$) was significantly lower than that of the C1-Polish group ($M = 12.55$, $SD = 8.79$). However, no significant difference was observed between the two C1 groups, with the non-Polish C1 group ($M = 12.24$, $SD = 8.95$) having a similar score to its Polish counterpart.
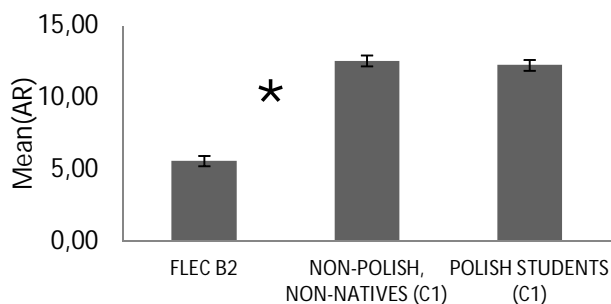

Figure 2 ANOVA POST-HOC results showing the MEAN CUMULATIVE distribution of the analyzed morphemes across groups

The data in Figure 2 presents the findings from Figure 1 in a more immediately accessible format. The Polish B2 learners taking a certificate examination at the University of Warsaw use the four morphemes with a significantly lower frequency than the C1 groups. Also, as already indicated for Figure 1 above, the differences in the linguistic background of C1 respondents (Polish vs. non-Polish) are largely negligible. This indicates an increase of learning with experience in the more advanced groups but offers no evidence for transfer at lower levels.
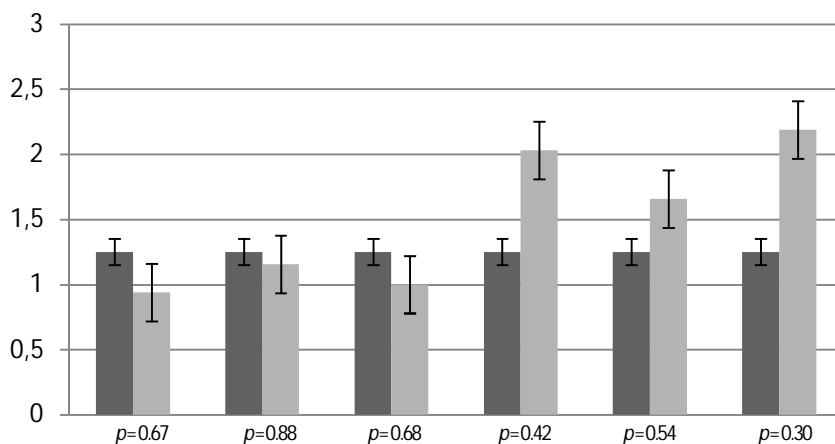

Figure 3 Unpaired *t*-test for the frequency of the Saxon genitive (L1 English vs. C1 English, L1 Polish)

To complement our primary analysis, we compared the frequencies of the analyzed morphemes between our learner corpora and the mini-corpus of native speaker output that was available to us. The L2 corpora (B1, C1 Polish, C1 non-Polish) were divided into smaller subsets (*N* = 32 each) to match the NS minicorpus

(*N*= 32). Both corpora (NS and NNS) were compared using unpaired (independent sample) *t*-tests. As is apparent in Figure 3, the use of the *'s* morpheme differs the least between all tested groups, including non-Polish C1 English users. We found that the discrepancies between the Polish and non-Polish C1 in the use of the Saxon Genitive ($p = 0.67$, $p = 0.88$, $p = 0.68$, $p = 0.42$, $p = 0.54$, $p = 0.30$) are not statistically significant. Implications about any kind of role of genitive-type transfer from Polish to English cannot be made with the results as they are; at the same time, however, there is no evidence against it; our results regarding the use of the possessive while not strictly negative, are inconclusive on all counts, as shown by the figures below ($N = 32$ for each tested group).

As seen in Figure 4, ($p = 0.78$, $p = 0.87$, $p = 0.32$, $p = 0.13$, $p = 0.6$, $p = 0.01$, $p = 0.96$, $p = 0.75$, $p = 0.25$, $p = 0.73$), with one anomalous outlier exception, the results of comparing L1-Polish C1 users with non-L1 Polish C1 users with respect to the *'s* morpheme remain inconclusive.
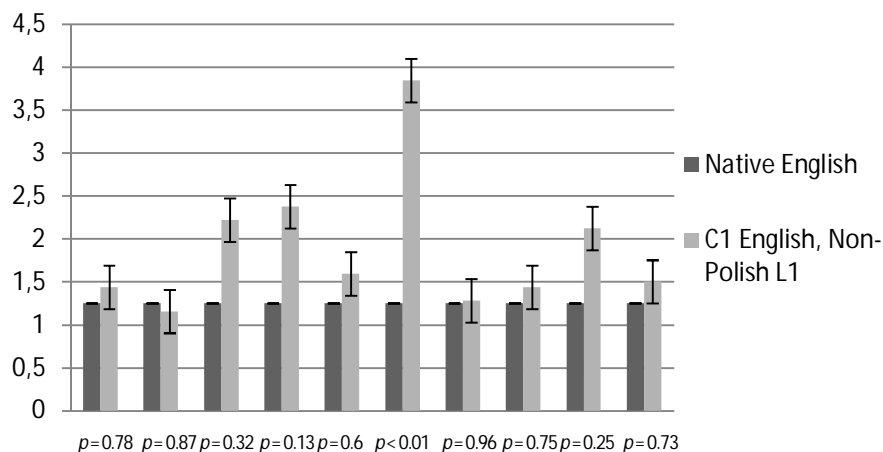

Figure 4 Unpaired *t*-test for the frequency of the Saxon genitive (L1 English vs. C1 English, L1 non-Polish)

The fact that the distribution of *'s* yields the same results for both the Polish and the non-Polish C1 groups when compared with the native controls offers direct support for the observation that transfer cannot be used to account for the distribution of the analytic genitive in the Polish B2 – C1 data. Figures 1 and 2 above seemed to tell the same story.

Assuming the negative correlation hypothesis to reflect the natural learning curve, we should expect the differences between the B2 and C1 learners to indicate straightforward gains in knowledge and skills, unrelated or weakly related to transfer. To appreciate the extent of the B2/C1 differences, consider Figure 5. As

far as the use of the analytical genitive morpheme is concerned, it is evident that there are pronounced frequency differences between the B2-English, L1-Polish group and the C1-English, L1-Polish group. A clear spike in frequency can be observed in more advanced learners. A series of *t*-tests conducted on matched subgroups of learners (N = 32/group) further confirms this finding (p< 0.01 for all tests). Thus the analytical genitive emerges as being *criterial* (Hawkins & Filipovič, 2012) for identifying CEFR levels of linguistic competence, a finding that should be of utmost importance to language teachers and educators.
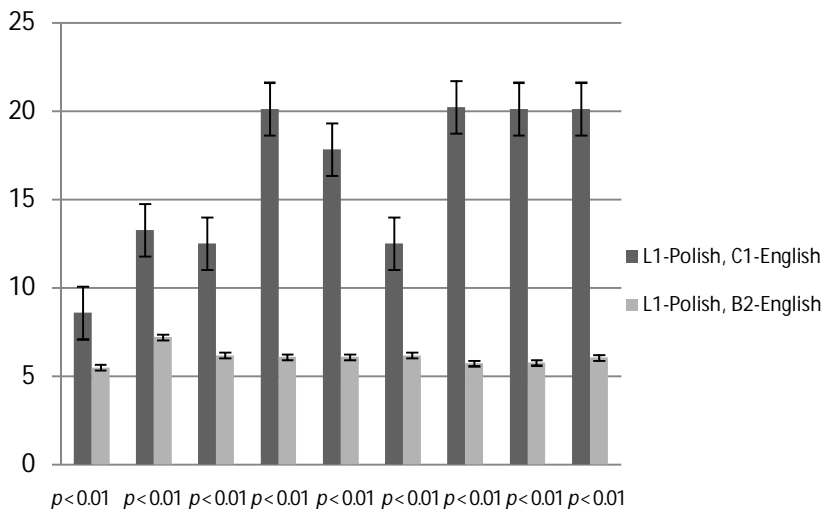

Figure 5 Successive *t*-test results between matched C1 and B2 groups with the Polish L1 background (analytical genitive)

Upon further comparison, we found that the frequency of the analytical genitive morpheme in the L1-Polish, C1-English groups (N = 32 each) is not significantly different from our modest (N = 32) native group (p = 0.14, p = 0.05, p = 0.14, p = 0.64, p = 0.71) with one exception (p < 0.01), as seen in Figure 6. In this case, the observed outlier test has to be attributed to chance and classified as a random occurrence. Thus, this confrontation of native and non-native data corroborates the findings of our initial ANOVA.

In regard to specific and non-specific determiners, we expected to find a significant difference between levels due to the fact that Polish makes no use of determiners and offers no cues to their use.
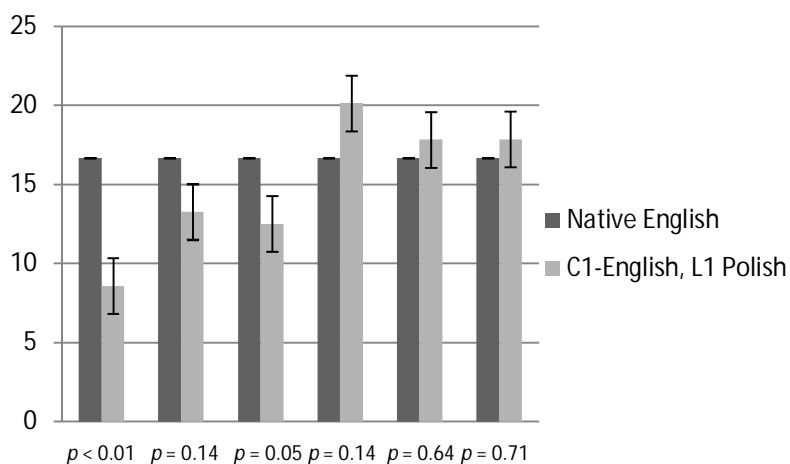
p < 0.01   *p* = 0.14   *p* = 0.05   *p* = 0.14   *p* = 0.64   *p* = 0.71

Figure 6 Frequency of the analytical genitive [of] in native-speaker output and advanced Polish-learner output

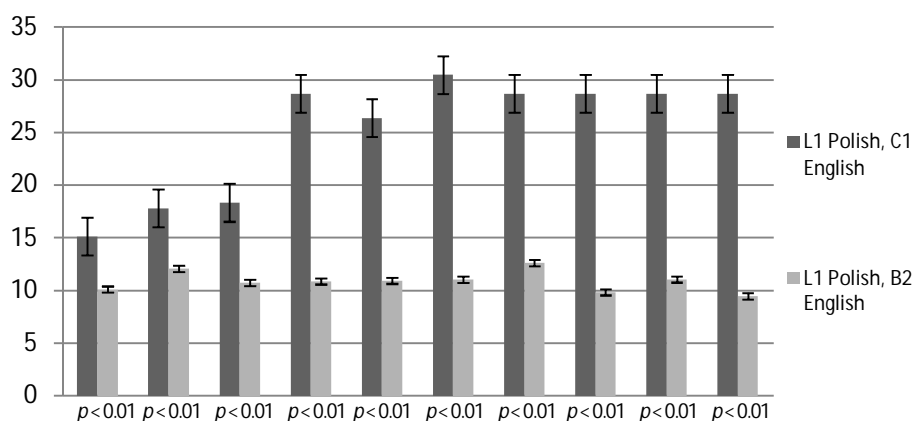*p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01 *p*<0.01

Figure 7 Definite article use in C1 and B2-level Polish learners of English

While we did find a significant difference ($p < 0.01$ for all tests performed) between the B2 and C1 groups, there was no difference in frequency between both C1 level groups, which again validates the claim that the difficulties in the use of the determiners in English at the post-intermediate stage (B2, C1,C2) are not unique to Polish learners and are not transfer-related.[6] These results were further supported by additional unpaired *t*-tests between subgroups (*N* = 32 each). Figure 7 illustrates the use of specific determiners as measured by our

---

[6] Transfer may still be operative at lower levels of linguistic competence, though. To shed more light on the issue, we intend to replicate the present study for CEFR levels A2 and B1.

test series. Again, a statistically significant difference ($p < 0.01$ for all tests performed) between levels can be seen here. The test yielded similar result in regard to non-specific determiner use, as shown in Figure 8.
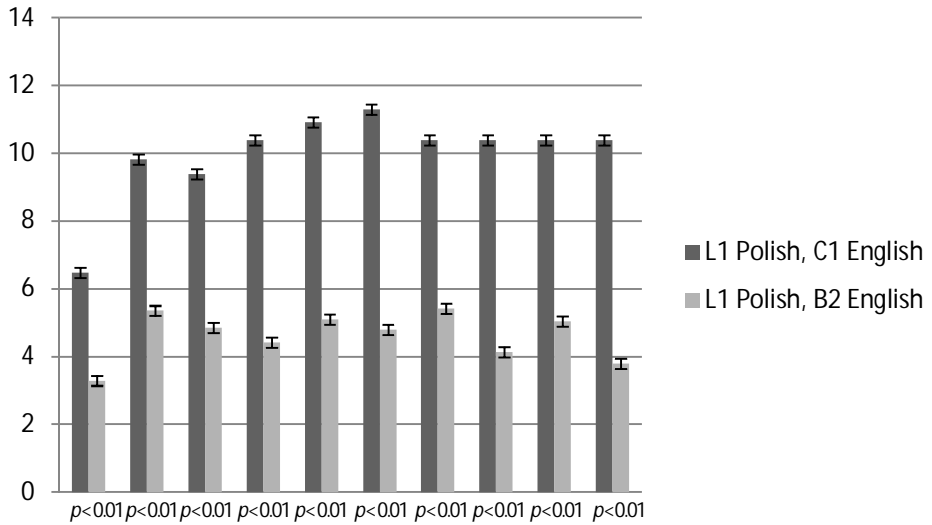

Figure 8 Indefinite article use (MEAN) in B2 and C1 Polish users of English

However, comparing our L1 Polish, C1 English groups (N = 32 each) with our native English group yielded inconclusive results concerning determiner use (Figures 9 and 10).
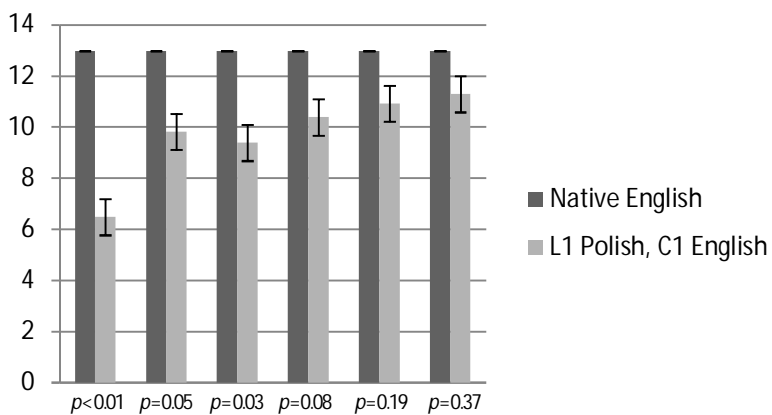

Figure 9 Indefinite article use (MEAN) in Native English and C1 Polish Learner English

A clear-cut statistically significant tendency cannot be pinpointed either for the non-specific (indefinite) article ($p < 0.01$, $p = 0.05$, $p = 0.03$, $p = 0.08$, $p = 0.19$, $p = 0.37$) or for the specific (definite) article ($p = 0.00$, $p = 0.00$, $p = 0.00$, $p < 0.01$, $p = 0.31$, $p = 0.16$, $p = 0.57$). We found these results puzzling but in-line with those pertaining to the use of the Saxon genitive. Both analyses show that more proficient learners may vary in how they use the morphemes under discussion. Still, larger amounts of data need to be accumulated and analyzed in order to confirm these findings.
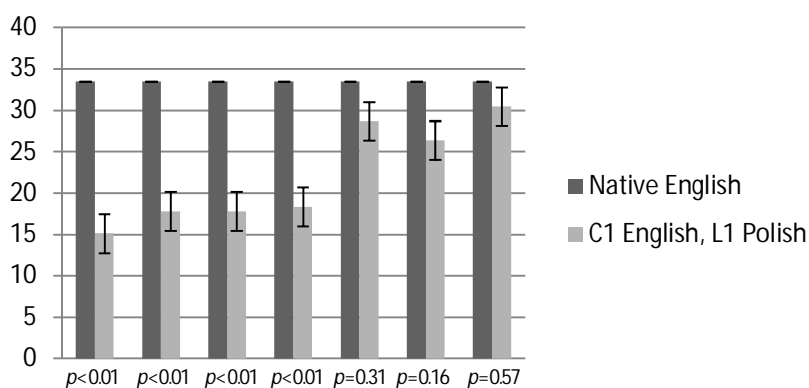

Figure 10 Definite article use in Native English and C1 Polish Learner English

In summary, our analyses primarily confirm a significant effect of learner advancement level on morpheme distribution. We found no significant effect of L1 background on morpheme use at the level of C1, as both Polish and non-Polish C1 learners reach frequencies that are similar to those found in native-speaker data. At the same time, our findings lend support to a developmental analysis of the four morphemes under investigation, based on the criteria of salience, type frequency and semantic scope (Collins et al., 2009). The developmental analysis correctly predicts the preference for the syntactically motivated *-of* over the inflectional *'s* and the preference for the definite article over the indefinite article. As for the problems with the use of English articles by Polish learners at more advanced levels, we did find marked improvement between levels B2 and C1 (Figure 8), which is entirely consistent with the developmental model, and the distribution of articles in non-native C1 written production, while systematically lower than in the reference native corpus (Figure 9) does not reach the level of statistical significance.

5. Discussion

The results of our study allow us to conclude that the L1 background of a learner group has a negligible effect on the frequency of the studied morphemes, that

is the possessive *'s*, the genitive *of*, and the definite and indefinite articles at CEFR levels B2 and C1.This, in turn, warrants the conclusion that transfer effects on morpheme frequency are marginal in comparison to the effects of general proficiency on the use of these morphemes. Naturally, whether this applies to all, most, or only selected morphemes remains to be determined, and there can be no doubt at this juncture that more studies are needed. Also, the distribution of the four grammar morphemes at lower levels (A2-B1) needs to be examined in the framework proposed here, to determine the degree of transfer among beginning and pre-intermediate students.

First off, we reject the hypothesis that the use of *the/a* and both genitive forms by Polish users at more advanced levels can be attributed to the influence of their native language. Our data clearly demonstrates that the features that we inspected exhibit similar distributional properties in the L2 English of both Polish and non-Polish groups. The early stages of article use are language-specific and appear both in target languages that have an article system and those which do not. For Polish learners, however, the early L1 influence manifests itself mostly as "lack of guidance" rather than specific interlanguage patterns. As for the more advanced stages, even proficient Polish learners exhibit statistically significant discrepancies in the use of the definite and indefinite articles, contrary to what may be expected, given the transfer-based account and the negative correlation hypothesis.

These patterns of distribution are not consistent with the developmental model of Collins et al. (2009) either – there are few semantic restrictions imposed on the type of nouns that articles are adjoined to, and in terms of distribution articles belong to the most frequently used English words. These facts point towards the alternative hypothesis that the distribution of articles is related to the general level of linguistic advancement. We have not, however, ruled out the possibility that articles in English represent a heavily marked, idiosyncratic feature, with little communicative impact and no full equivalent in any of the languages used in the corpus data for the purposes of this paper, so that even in languages with an overt article system the overlap with English is minimal or partial at best. Should this turn out to be the case, a transfer-based account would still be possible, as every one of the languages under investigation would have its own, unique, ways of encoding (or ignoring) the information expressed by the articles in English.

Secondly, addressing our "genitive" research questions, the B2-level Polish group did not overuse the inflectional possessive, invalidating the predictions made by contrastive analysis and again demonstrating a lack of transfer effects, despite inflection being the only way to signal grammatical relations in Polish.

On the one hand, it may point to a significant role of developmental factors, encouraging the use of the periphrastic genitive, matching the preferred Polish word order for the genitive construction. On the other hand, transfer may still be operative at earlier stages of linguistic sophistications (CEFR levels A1, A2, B1), being gradually replaced with proceduralized target knowledge on a regular learning curve at level B2. This analysis is consistent with the observed increased variability in the use of *'s* at C1, irrespective of the linguistic background of the learners. This strongly suggests that even advanced users may acquire certain structures at markedly different paces and use them to different degrees.

Our results allow us to conclude that the L2 patterns of the *'s* and *of* distribution do not appear to be L1-specific in monitored use. This is to say that we did not find any notable differences between Polish and non-Polish L2 speakers of English in this respect. What is more, the difference between proficiency levels (B2 vs. C1) far outweighs any minute (and statistically insignificant) discrepancies that can be pinpointed between the C1 (Polish and non-Polish) groups. For this reason, we reject the hypothesis that the patterns of genitive and possessive distribution are unique to Polish learners and attributable to a strong transfer effect. Accordingly, we consider viable the hypothesis of identical development (i.e., that learners from different backgrounds follow roughly the same developmental stages in the use of the two markers).

Furthermore, we did observe some evidence of leveling off between native speakers and advanced learners. As learners become more advanced, the frequency of selected structures begins to even out across native and non-native corpora. These findings (in particular the lack of language-specific patterns in the distribution of genitive/possessive markers) suggest that the negative correlation hypothesis need not be the preferred account of the observed tendency for the NNS use of the two morphemes to eventually resemble their NS distribution. Alternatively, the leveling-out could be due to training/learning effects, prolonged exposure, frequency of use, pedagogical intervention in the classroom or numerous other non-transfer related causes. However, given the nature of our findings (Figures 3-4, and 9-10), the discrepancies between L1 and L2 usage remain big enough to question the legitimacy of defining foreign language competence with reference to native or native-like grammaticality standards.

6. Conclusions

In this paper, we have attempted to address some basic questions pertaining to transfer effects in the area of filler morphemes (in the sense of Mosteller & Wallace, 1963), or function words, in Polish speakers of L2 English through the use of corpus linguistics. We chose this methodology to seek out evidence of transfer

without resorting to error analysis. Our objective was to study the occurrence of transfer effects at two proficiency stages (B2 and C1) and between two L1 backgrounds (Polish and non-Polish) with additional reference being made to a small control corpus of native English (for the purpose of disentangling transfer effects from L2 developmental features).

The most significant observation afforded by our analyses is that the frequency of occurrence of the possessive/genitive morphemes can be correlated primarily with developmental factors, while any effect of L1 background remains marginal. Counter to the transfer-based predictions, we found that the use of the possessive marker in the English of Polish learners does not decrease in frequency between advancement levels. We found that the frequency of these markers is higher in the C1 group ($M = 1.49$, $SE = 0.18$) when compared to the B2 group ($M = 0.82$, $SE = 0.18$). Analogous observations can be made about both articles, with the indefinite article occurring much less frequently in the B2 group ($M = 4.70$, $SE = 1.13$) that in either of the C1 groups ($M = 11.23$, $SE = 1.13$; $M = 9.70$, $SE = 1.13$). These finding are consistent with the criteria of salience and frequency, in the sense of Collins et al. (2009).

Our findings go beyond questioning the role of transfer in the acquisition of L2 grammar morphemes and lending support to developmental accounts. We also hope to have demonstrated that stylometric insights have the potential of being directly applicable to corpus-based analyses of linguistic competence. We remain hopeful that this research perspective will be promulgated, offering much needed quantitative solutions to the many unresolved questions of SLA studies. The proposed model has an extra advantage of being adaptable to suit classroom needs, as a distribution-based analysis avoids any excessive focus on errors. Given the anti-error methodological bias of today's language teaching methodology, the difficulties with identifying native standards for reference purposes, the preoccupation with communicative effectiveness at the cost of formal accuracy, the emergence of functional competence intended to replace native(-like) competence, the painstaking, time-consuming and subjective process of error mark-up, we believe that a distribution-based model has a chance of becoming an important pedagogical tool in language education.

References

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

Anthony, L. (2014b). AntConc (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Arabski, J. (1979). *Errors as indications of the development of interlanguage.* Katowice: Uniwersytet Śląski.

Bardovi-Harlig, K. (2001). Evaluating the empirical evidence: Grounds for instruction in pragmatics? In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp 13-32). Cambridge: Cambridge University Press.

Belz, J. A., & Vyatkina, N. (2008). The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning and Technology, 12*(3), 33-52.

Benati, A. (2001). A comparative study of the effects of processing instruction and output- based instruction on the acquisition of the Italian future tense. *Language Teaching Research, 5*(2), 95-127. dx.doi.org/10.1177/13621688010 0500202

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning- output instruction on the acquisition of the English past simple tense. *Language Teaching Research, 9*(1), 67-93. http://dx.doi.org/10.1191/ 1362168805lr154oa

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*(1), 1-17.

Brown, R. (1973). *A first language: The early stages.* London: George Allen & Unwin.

Bu, J. (2012). A study of relationships between L1 pragmatic transfer and L2 proficiency. *English Language Teaching, 5*(1), 32-43.

Cancino, H., & Rosansky, E. J., & Schumann, J. H. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 207-230). Rowley, MA: Newbury House Publishers, Inc.

Collins, L., Trofimovich, P., White, J., Horst, M., & Cardoso, W. (2009). Some input on the easy/difficult grammar question. *Modern Language Journal, 93*, 336-53.

Corder, S. P. (1967). The significance of learners' errors. International *Review of Applied Linguistics, 5*(4), 161-170.

Davies, M. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc.

Davies, M. (2008-) *The corpus of contemporary American English (COCA): 520 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R., & Barkuizen, G. (2005). *Analyzing learner language.* Oxford: Oxford University Press.

Ekiert, M. (2004). Acquisition of the English article system by speakers of Polish in ESL and EFL settings. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics, 4*(1). Retrieved from https://tesolal. columbia.edu/article/acquisition-of-english-article

Faerch, C., & Kasper, G. (1989). Internal and external modification in interlanguage request realization. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 221-247). Norwood: Ablex.

Gelbukh, A., & Sidorov, G. (2001). Zipf and Heaps Laws' Coefficients depend on language. Proc. CICLing-2001, Conference on Intelligent Text Processing and Computational Linguistics, February 18-24, 2001, Mexico City. Lecture Notes in Computer Science N 2004, ISSN 0302-9743, ISBN 3-540-41687-0, Springer-Verlag: 332-335.

Hawkins, J. A., & Filipovič, L. (2012). *Criterial features in L2 English. Specifying the reference levels of the Common European framework.* Cambridge, Cambridge University Press.

Holliday, A. R., Aboshiha, P., & Swan, A. (Eds.). (2015). *(En)countering native-speakerism: Global perspectives.* London: Palgrave Macmillan.

Hussein, A. A., & Mohammad, F. M. (2011). Negative L1 impact on L2 writing. *International Journal of Humanities and Social Science, 1*(18), 184-195.

Ionin, T., Zubizarreta, M. L., & Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua, 118,* 554-576.

James, C. (1980). *Contrastive analysis.* Essex: Longman

Jarvis, S., & Odlin, T. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language learning, 50*(2), 245-309.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition.* New York: Routledge.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly 40*(1), 157-181.

Jenkins, J. (2007). *English as a lingua franca: attitude and identity.* Oxford: Oxford University Press.

Johansson, S. (1998a). On the role of corpora in cross-linguistic research. In S. Johansson & S. Oksefjell (Eds.), *Corpora and cross-linguistic research: Theory, method, and case studies* (pp. 3-24). Amsterdam and Atlanta, GA: Rodopi.

Johansson, S. (1998b). Loving and hating in English and Norwegian: A corpus-based contrastive study. In D. Albrechtsen, B. Henriksen, I. M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy.*

Essays presented to Kirsten Haastrup on the occasion of her sixtieth birthday (pp. 93-103). Odense: Odense University Press.

Kachru, B. (1985). Standards, codification, and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.

Kellerman, E. (1983). Now you see it, now you don't. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 112-134). Rowley, MA: Newbury House.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* University of Michigan Press: Ann Arbor.

Lee, J., & VanPatten, B. (2003). *Making communicative language teaching happen.* New York: McGraw-Hill.

Leech, G. (1994). Students' grammar – teachers' grammar – learners' grammar. In M. Bygate, A. Tonkyn, & E. Williams (Eds.), *Grammar and the language teacher* (pp. 17-30). London: Prentice Hall International.

Lewis, M. (1993). *The Lexical Approach. The state of ELT and a way forward.* Hove: Language Teaching Publications.

Lewis, M. (1997). *Implementing the Lexical Approach.* Hove: Language Teaching Publications.

Lightbown, P. M. (1984). The relationship between theory and method in second language acquisition research. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 241-252). Edinburgh: Edinburgh University Press.

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative teaching: Effects on second language learning. *Studies in Second Language Acquisition, 12*(4), 429-448.

MacKenzie, I. (2014). *English as a lingua franca: Theorizing and teaching English.* London: Routledge.

Master, P. (1987). A cross-linguistic interlanguage analysis of the acquisition of the English article system. Unpublished doctoral dissertation, University of California at Los Angeles.

Mayo, G., & Pilar, M. (2008). The acquisition of four non-generic use of article "the" by Spanish EFL learners. *System, 36,* 550-565.

Mollin, S. (2006). *Euro-English: Assessing variety status.* Tübingen: Gunter Narr.

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association, 58,* 275-309.

Otwinowska-Kasztelanic, A. (2016). *Cognate vocabulary in language acquisition and use. Attitudes, awareness, activation.* Bristol: Multilingual Matters.

Phillipson, R. (1992). *Linguistic imperialism.* Oxford: Oxford University Press.

Phillipson, R. (2008). Lingua franca or lingua frankensteinia? English in European integration and globalization. *World Englishes, 27*(2), 250-267.

Prator, C. (1967). *Hierarchy of difficulty: Unpublished classroom lecture.* University of California, Los Angles.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning.* Clevedon: Multilingual Matters.

Schachter, J. (1974). An error in error analysis. *Language Learning, 24*(2), 205-214.

Seidlhofer, B. (2006). English as a lingua franca in the expanding circle: What it isn't. In R. Rubdy & M. Saraceni (Eds.), *English in the world* (pp. 40-50). London: Continuum.

Swan, M., & Smith, B. (ed.) (2001). *Learner English: A teacher's guide to interference and other problems.* Cambridge: Cambridge University Press.

Śpiewak, G., & Gołębiowska, L. (2001). Polish speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 162-178). Cambridge: Cambridge University Press.

Stockwell, R. P., Bowen, J. D., & Martin, J. W. (1965). *The grammatical structures of English and Spanish.* Chicago: University of Chicago Press.

Thorne, S., & Lantolf, J. (2004). Center for advanced language proficiency education and research. Invited address. Inter-Agency Language Roundtable, Washington D.C., March 18, 2004.

Thorne, S. L., & Payne, J. S. (2005). Evolutionary trajectories, Internet-mediated Expression, and language education. *CALICO Journal, 22*(3), 371-397.

Thorne, S., & Payne, S. (2006). Corpus linguistics and language development: Research, assessment, and pedagogical innovation. Invited address. University of California, Davis, April 26, 2006.

Treffers-Daller, J. (2011) Operationalising and measuring language dominance. International *Journal of Bilingualism, 15*(2), 147-163.

Van Patten, B. (1996). *Input processing and grammar instruction.* Norwood, NJ: Ablex.

VanPatten, B. (2003). *From input to output. A teacher's guide to second language acquisition.* Boston: McGraw Hill.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15,* 225-243.

VanPatten, B., & Oikkenon, S. (1996). Explanation vs. structured input in processing instruction. *Studies in Second Language Acquisition, 18,* 495-510.

Wang, Y. (2001). L2 Learning of the English passive. Unpublished doctoral dissertation. Guangzhou: Guangdong University of Foreign Studies.

Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly, 28(2),* 377-89.

Widdowson, H. G. (1997). EIL, ESL, EFL: Global issues and local interests. *World Englishes, 16*(1), 146-53.

Wishnoff, J. (2000). Hedging your bets: L2 learners' acquisition of pragmatic devices in academic writing and computer-mediated discourse. *Second Language*

Studies: Working Papers of the Department of Second Language Studies, University of Hawai'i, 19, 119-157.

Wode, H., Bahns, J., Bedley, H., & Frank, W. (1978). Developmental sequence: an alternative approach to morpheme order. *Language Learning, 28*, 175-185.

Zipf, G. K. (1935). *The psychobiology of language.* Houghton-Mifflin.

Zobl, H. (1980). The formal and developmental selectivity of L1 influence on L2 acquisition. *Language Learning, 30*(1), 43-57.