Word-formation preferences of non-natives

Richard Skultéty Madsen, Section of English Business Communication, Department of Culture and Global Studies, Aalborg University, Denmark

This paper presents the preliminary results of a study of the word-formation preferences in their second/foreign languages of non-native speakers. The study attempts to bring together linguistic typology and second language acquisition. Several pairs of native – non-native languages, for example Danish (NL) – German (NNL), Slovene (NL) – English (NNL), were tested. The informants are university students of the respective foreign languages. It was investigated which word-formation processes (premodification, postmodification, or combinations hereof) are preferred by the informants by measuring which word-formation processes come to the informants' minds first when asked to produce new words from certain keywords. The preferences of word-formation processes were correlated both with the informants' native languages and the target language. The purpose of the correlation was to examine whether the word-formation choices of the informants are influenced more by the typological nature of their mother tongues or by the typological nature of the target languages. The differences between the informant groups could not be attributed to their mother tongues unequivocally. Instead, general intergroup tendencies were found with respect to preferences in word formation.

Keywords: *word formation*, *cross-linguistic influence*, *second language acquisition*, *linguistic typology*, *statistical analysis*

1. Introduction

The purpose of this paper is to investigate whether there are differences between various groups of learners of English as a second language with respect to their preferences of forming words in English as a function of their first languages (L1). This study is thus on the crossroads between linguistic typology and second language acquisition although the main focus is on the production of the acquired second language, not on its acquisition per se. The use of the second language is tested in laboratory in a psycholinguistic experiment, not investigated in corpora. Hence, the informants' preferences are measured in terms of what new word forms they provide first in the experiment, not in terms of how frequently and in what contexts they use certain word forms.

At the time of writing this article, usable data were only available from three groups of informants, Danish, Israeli and Slovene university students, all partaking in the experiment with English as L2. Some data were also collected from informants with Danish and Slovene L1 and German L2. However, the amount of these data is unfortunately miniscule and does not make possible a comparison. It is planned to pursue the collection of further data in the future both with respect to L1's and L2's.

2. Theory and hypotheses

2.1 *Theory*

The main theory of this study is the theory of cross-linguistic influence, i.e. one's mother tongue (L1) – and likely other languages one has acquired – influences the acquisition of a new language (Lado 1957; Johnson 2008; Jarvis 2011). It has also been demonstrated that languages acquired more recently can retroactively influence the languages one has acquired earlier (Pavlenko & Jarvis 2002; Jarvis & Pavlenko 2008; Madsen 2015), but this possibility is not explored in this study.

The main hypothesis is thus a fairly general one: There is a statistically detectable difference between how groups of learners with different mother tongues prefer to make up new words in their target language. To test this hypothesis, informants with typologically different mother tongues, currently Danish, Hebrew and Slovene, have been enlisted to provide data on their use of English as their L2.

A hypothesis based on concrete differences between the informants' mother tongues is that the Danish informants will use compounding more than the other groups. This hypothesis is based on the nature of Danish grammar, or perhaps rather Danish pragmatics (Togeby 2003), preferring compounds to adjectival or prepositional paraphrases. For example, *universitetsprofessor* (university professor) in Danish is *univerzitetni professor* (university-y professor) or *professor na univerzi* (professor at university) in Slovene (Svane 1958; deBray 1980; Melita Koletnik, personal communication).

Apart from the above-mentioned hypotheses based on the theory of cross-linguistic influence, the opportunity was taken to investigate a couple of other hypotheses as well because the data collected for this project are also suitable for testing hypotheses below. One of these hypotheses is that the informants will prefer suffixes to prefixes. This has been posed as a general cross-linguistic tendency and has as such nothing to do with cross-linguistic influence in language acquisition (St. Clair et al. 2009; Ramscar 2013).

Another hypothesis concerns the assumption that the informants, when asked to make new words, will prefer making new lexemes instead of *just* providing different forms of the same lexeme expressed by the keyword. In other words, the informants will prefer using derivation and compounding to using inflection. The justification for this hypothesis is an informal observation that laymen tend to equate words with lexemes, and thus if their task is to make new words, they will tend to provide new lexemes based on the keyword (i.e. use derivation and/or compounding) instead of giving various inflectional forms of the keyword.

Here is an overview of the hypotheses to be tested:

1) There is a statistically significant difference between the informant groups' preferences with respect to word formation.
2) The Danish informants use compounding more than the Israeli and Slovene informants.
3) All the informant groups use postmodification more than premodification.
4) All the informant groups use derivation/compounding more than inflection when producing new "words".

In the rest of this section, major theoretical challenges are discussed briefly.

2.2 *Affix or root?*

The original idea was to measure the differences between the informant groups in terms of their use of derivation by way of affixation and compounding, i.e. the combining of roots. However, it was soon to become apparent that the demarcation of roots from affixes is not at all clear-cut (Štekauer et al. 2012). Especially, deciding the status of a number of morphemes as being prefixes or roots proved to be difficult. These morphemes include among others *out*, *over*, *under*, *geo*, *hyper*, *super*, *uni*.

Looking at the matter superficially, it may be preferred to consider the elements of Germanic origin (*out*, *over*, *under*, etc.) to be roots inasmuch they can appear as orthographic words. In contradistinction, the elements of Latin/Greek origin may be preferred as prefixes since they do not ordinarily constitute orthographic words. Lexicographers do not seem to agree. For instance, American Heritage Dictionary (2018) defines *geo* a prefix whereas Random House Kernerman Webster's College Dictionary (2018) calls it a combining form, i.e. a root. Some of the morphemes above can in fact appear as orthographic words as abbreviations of words in which they would otherwise be prefixes, for instance *hyper(active)*, *homo(sexual)*, *trans(sexual)*. Additionally, Bauer (2017) notes that these morphemes are sometimes written as orthographic words as though they were treated as adjectives by the language users.

Since it is beyond the scope of this paper to try to settle the matter of affix vs root once and for all, a more practical and hopefully less controversial approach has been adopted. Instead of counting the occurrence of prefixes, suffixes and roots, modification has been used as the metric for measuring the differences between the informant groups. Three categories of modification have been distinguished: (1) the keywords of the questionnaire being premodified, (2) the keywords being postmodified, (3) the keywords being used as premodifiers. In the first category fall prefixes (e.g. ***re****volve*) and roots preceding the keyword (e.g. ***arm****chair*). In the second category fall suffixes (e.g. *vision****ary***) and verbal particles (e.g. *get* ***away***). The third category comprises keywords modifying a following element in a compound (e.g. ***chair****man*).

In this way, the distinction between prefixes and roots becomes irrelevant since both prefixes and roots placed before other roots are premodifiers. Thus, it does not matter whether *over* is considered a prefix or a root because it is a premodifier in say *overworked* regardless of its morphological status. However, it is a different matter for the distinction between suffixes and roots because a suffix postmodifies the element that precedes it, and a root is premodified by the element that precedes it. This was a matter to consider for, among other elements, *log(y)* (stemming from Greek *logos* and frequently appearing in names of branches of science). On etymological grounds, it should be considered a root; however, one might also argue that it is a suffix in modern English on a par with *-ism*. Ultimately, it was decided to count as a root following the reasoning in Merriam-Webster (2018), considering it a combining form. The same applies to the element *phobe*.

A special case is that of the keyword *trans*. When it was used in its canonical prefixal (American Heritage Dictionary 2018) sense (e.g. in *transaction*), it was counted as a premodifier, not as a premodifying keyword. When it was used in its sense of 'transgender' (e.g. *transphobic*) it was counted as a root, and thus as a premodifying keyword.

Finally, there were few responses that could not be analyzed in terms of modifications, either. These are dealt with separately in the analysis.

2.3 *To be or not to be analyzed?*

Testing the hypotheses posited requires a morphological analysis of the responses so that it can be ascertained whether the informants use prefixes, suffixes or roots, or as discussed above premodification or postmodification. In certain cases, however, it is not clear whether a part of a response should indeed be analyzed or not. It is for instance the case when words such as *nature*, *nation* or *aggressive* appear in compounds provided by the informants. These words are clearly complex morphologically; however, it is not given that the informants are or can be expected to be aware of the complexity since these words have been borrowed from Latin and formed according to the morphotactic properties and rules of Latin, which people without classical training are not likely to know. There is no reason to assume that a significant number of the informants know Latin.

A decision as to whether to decompose members of compounds was made on word by word basis. Words such as *nature* and *nation* were not decomposed because although their endings should be recognizable by the informants, their root is not likely familiar since it does not appear in modern English freely, and for laymen it is, therefore, not conspicuous that these words are actually related. On the other hand, *aggressive* and similar words were decomposed because even though its root is not used freely in modern English, it should be recognizable since it appears in a number of other words, e.g. *progressive*, *aggressor*, *regression*, which even laymen with a little fantasy can diagnose as related (having to do with movement).

A similar situation arises when the informants use suffixes such as *-ation*, which are originally a combination of several suffixes (four in the case of *-a-t-io-n*). Should the use of *-ation* and similar complex suffixes count as adding one or several suffixes? Since, as discussed above, it cannot be expected of the informants to know Latin, it was decided that such suffixes count as one.

2.4 *Inflection vs derivation*

As described above, one of the hypotheses concerns the expectation that the informants will use derivation and compounding more than inflection to create new words from the keywords given. However, it has transpired that it is not an easy matter to decide whether a response contains inflectional or derivational elements. Beyond the theoretically challenging question as to whether participles should be considered inflectional or derivational forms, homography causes problems too. For instance, without context, it is impossible to know whether the informants meant *thought* as a noun (derivation) or a past tense form (inflection), not to mention the possible participial reading as well. Unfortunately, the design of the investigation, namely the lack of context in the responses, does not make it possible to solve this issue. Therefore, only forms unambiguously inflectional were counted as such, that is the plural of nouns, 3sg present of verbs and the superlative of adjectives. In the case of keywords ambiguous between an adjective and a verb reading (e.g. *clean*), the *-er* suffix was counted as the comparative, thus inflectional, giving inflection the benefit of the doubt.

2.5 *How not to bias the informants with respect to part of speech?*

There was a challenge with the German version of the questionnaire. It was – just as the English version (see below in §3) – to represent all declinable parts of speech. The English

version intentionally contains keywords which are ambiguous between parts of speech, for instance *work*. The German version was to contain similarly ambiguous words as well. However, as is well known, German standard orthography requires that nouns be capitalized; consequently, potential ambiguity is reduced considerably. The solution was to write all keywords without capitalization and write a note for the informants that it was done on purpose.

3. Method

3.1 *Questionnaire design*

As the method of data collection, this study has used questionnaires administered electronically either through Moodle (for the Danish informants) or through Google Documents (for the Israeli and Slovene informants). The informants were asked to make as many words as they could based on the 22 keywords given. The informants were given no instruction as to what method of word formation to use; they were free to use whatever method they saw fit. However, they were not allowed to use aids such as dictionaries, and there was a time limit of 20 minutes for making new words of all the keywords. These limitations were strictly enforced for the Danish informants as they were under surveillance during the filling in of the questionnaires. However, it is not possible to ascertain whether the other informant groups kept the rules of the game or not.

Naturally, all the informants were given the same set of keywords to build on. However, whereas the Israeli and Slovene informants were given all the keywords at the same time in the same questionnaire, the Danish informants were given the keywords in two mutually exclusive half-subsets. The reason for this was that the Danish informants were given the questionnaire in class in order to secure their proper participation in the project; the time used on the questionnaire was, however, to be limited to 2 x 10 minutes so that the filling in the questionnaire would not be detrimental to the curriculum. On the other hand, the Israeli and Slovene informants filled in the questionnaire outside the classroom, which can explain their relatively low participation.

The keywords given in the questionnaire are the following: ***vision***, *persuade*, ***passive***, *pose*, *human*, ***evolve***, *chair*, *girl*, *world*, *high*, *graph*, *use*, *form*, *think*, *green*, *scribe*, *help*, *work*, *get*, ***trans***, *clean*, *nerve.* Both Moodle and Google Documents were programmed to present the keywords to the informants in a random order (Oppenheim 1992; Dörnyei 2014).

An attempt was made to represent all the inflectional parts of speech (nouns, verbs and adjectives) equally and both roots of Germanic and roots of Latin/Greek origin. Except for the highlighted items, all the keywords are monomorphemic roots. The polymorphemic keywords were included in order to test whether the informants were prepared to separate the morphemes from one another in their formation of new words. *Trans* was included because it can be both a prefix (or premodifier cf. §2.2) and a root (as the abbreviation of *transsexual*), and it was thus tested which use the informants would prefer.

3.2 *Preparation of the responses for analysis*

The preparation of the responses for the analysis was an unexpectedly lengthy process, involving several steps. As it turned out, several informants did not use any conventional

separators (e.g. comma) to separate their responses to the same keyword from each other. Nor were the informants as a group consistent in using UK vs US spelling, not to mention plain misspellings.

Consequently, the responses had to be read through by a human in order to detect whether a sequence of words in the response of the informants to a certain keyword was meant as a compound or separate responses. When compounds were detected, hyphens and spaces between the elements of the compounds were ignored because since the informants did not write texts, it was impossible to know whether their orthographical practice had any significance meaning-wise.

At the same time as detecting compounds, all responses were converted to lowercase, and clearly incorrect responses, e.g. the providing of synonyms of the keyword instead of derivations/compounds based on the keyword, were discarded. Misspelled responses were marked for later reanalysis. All responses were standardized into US spelling in order to ease automated analysis later on.

Additionally, unconventional but in principle possible responses, e.g. *subvision* or *impersuade*, were annotated as such; however, later on in the analysis, they were treated just as ordinary responses, i.e. words which are attested in a dictionary. On the other hand, responses that violate morphotactic rules of English derivation were eliminated from the database, e.g. *worldness* because *-ness* is not ordinarily added to nouns, but to adjectives, and *world* is not attested as an adjective (Huddleston & Pullum 2002).[1]

In the next step, misspelled responses were normalized into their standard orthography if and only if the intended word could be recovered beyond reasonable doubt. For instance, *usefull* was rejected as it could equally be a misspelling of both *useful* and *usefully* whereas say *visionery* was transformed into *visionary* and later analyzed as that word form.

In the final step of raw data preparation, the remaining responses were partitioned into their constituting morphemes. This step was done largely automatically with the use of a custom-made morphological analyzer in Microsoft Excel (Madsen unpublished). Manual intervention was only performed when needed. The morphological analyzer produced output consisting of prefixes, roots and suffixes in accordance with the original idea of using affixes and roots, i.e. derivation and compounding as metrics. This output was then converted into elements of modification in accordance with the description in §2.2 above. Also, responses not containing modifying elements were tagged as such.

Even though many of the keywords can belong to several parts of speech, a response was only counted as conversion if the informant marked it as such unequivocally, e.g. in the case of the keyword *form* as *to form* or *a form*. If the keyword was merely repeated in a response, as was the case with many responses of the Danish informants, it was considered an invalid response.

Apart from dividing the responses into their constituent morphemes, it was noted whether deletion was employed. It is relevant for the polymorphemic keywords, for which

[1] Admittedly, a clear-cut distinction between the accepted and not accepted unconventional responses may be difficult to draw. Since the informants' intention with the responses, i.e. what they thought the responses meant, was unavailable, the decision whether to accept one response and reject another one was based mainly on morphotactic rules presented in grammar books and not in dictionaries and corpus searches. Indeed, one of the reasons for conducting this research was to analyze students' creativity. In fact, it turned out in some cases that unconventional responses judged acceptable had been used by people as for instance proper names: Subvision, a music band.

deletion entails the removal of one of the elements and then adding new elements, for instance when producing *visually* from *vision*.

4. Analysis

4.1 *General comparison*

To start with, Table 1 presents some basic statistics of the data used.

Table 1: Basic statistics of the database

	DNK	**SLO**	**ISR**
number of informants	69	15	15
raw responses in orthographic words	4447	1220	1309
responses including compounds	4326	1197	1248
responses/informant	62.7	79.8	83.2
deletions	27	*11*	*21*
typos	167	41	35
recovered typos	108	28	26
well-formed but unattested words	68	8	10
unrelated words	127	60	10
incorrect derivations	41	11	6
valid responses	3898	1096	1225
valid responses/informant	56.5	73.1	81.7
valid responses/informant/keyword	2.7	3.3	3.7

As can be seen, the Israeli informants were the most productive in making new words. The following two figures show the differences between the informant groups for each of the keywords in terms of both tokens and types in the responses.

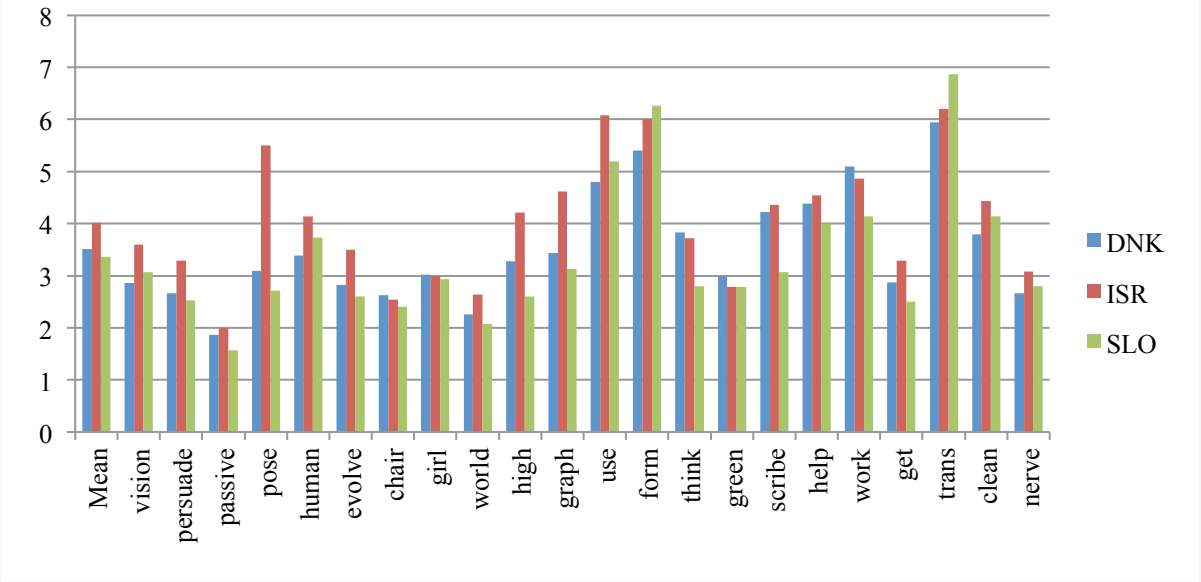

Figure 1: The informants' responses by token

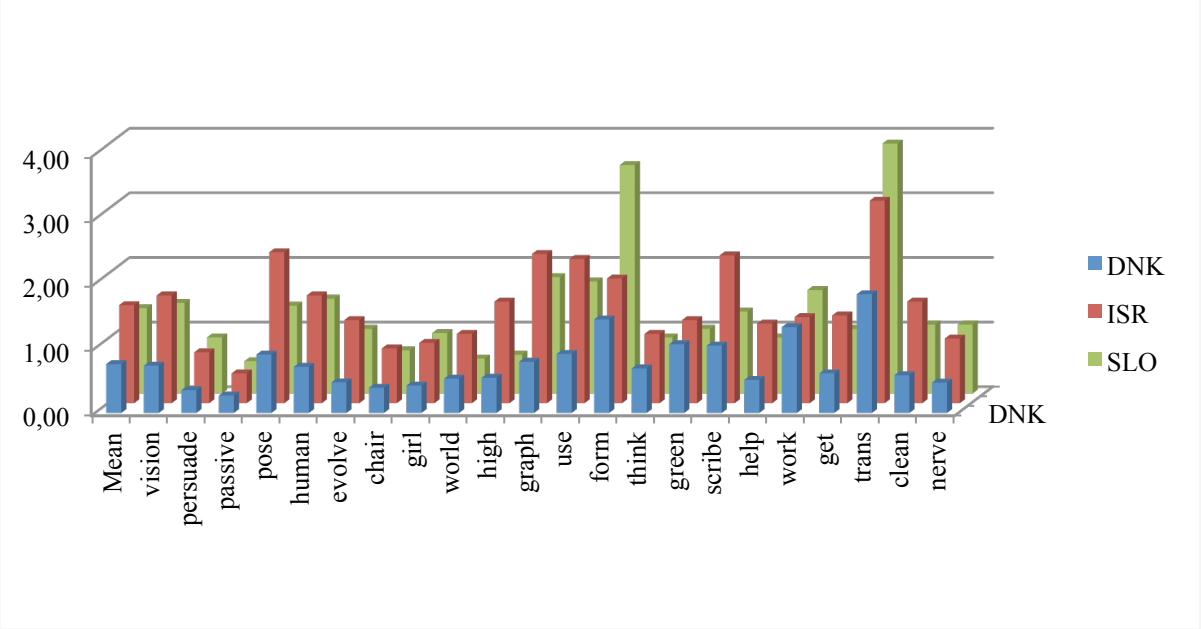

Figure 2: The informants' responses by type

The difference between the informant groups is significant when it comes to type. The Danish informants show significantly lower variety in their responses than the Israeli and Slovene informants. Table 2 shows the differences and their statistical significance as estimated by the two-tailed heteroscedastic t-test (Harmon 2014). Since multiple t-tests are used instead of ANOVA, the significance level of alpha is adjusted down to 0.017 from the usual 0.050 (Abdi 2007). This alpha is used in all the statistical estimations.

Table 2: The differences between the groups in terms of number of responses

	By token			By type		
	Means		p	Means		p
DNK vs ISR	3.52	4.02	0.149	0.75	1.52	3.29E-5
DNK vs SLO	3.52	3.36	0.668	0.75	1.32	8.70E-3
ISR vs SLO	4.02	3.36	0.090	1.52	1.32	0.389

The reason why the mean values by token by keyword in Table 1 and Table 2 differ from each other is that in Table 1, the mean was calculated by the number of all responses per group divided by the number of informants. However, not all informants attempted to make new words from all the given keywords. Hence, the averages in Table 2 have been calculated by only taking those informants into account who actually wrote a response to a given keyword. Thus, the divisor differs from keyword to keyword. The two different ways of calculating the mean for a group of informants are as follows:

In Table 1, the number of all responses (tokens) for all keywords combined is divided by the number of all the informants in the group. In Table 2, the number of responses (tokens and types, respectively) to an individual keyword is first divided by the number of informants who responded to that keyword. Then, these means are summed, and their sum is subsequently divided by the number of keywords. The same method is used in all the calculations below, except that premodifying elements, postmodifying elements and keywords used as premodifiers, respectively, are substituted for responses in the formula.

Despite the differences, which in two cases are statistically significant, it is interesting to note that the distribution of the number of responses per keyword is very similar for all three informant groups as can be seen from the graphs in Figure 1 and Figure 2. The informants seem to have found the same keywords more or less challenging. This is also shown in Table 3 below. Four of the five most inspiring keywords are the same for all three informant groups, and three of the five most tedious keywords are also the same.

Table 3: Most and least inspiring keywords

	Top 5						Bottom 5				
DNK	trans	form	work	use	help	…	persuade	nerve	chair	world	passive
ISR	trans	use	form	pose	work	…	girl	green	world	chair	passive
SLO	trans	form	use	work	clean	…	persuade	get	chair	world	passive

Another rough measure of the difference between the groups is the complexity of their responses. Complexity is measured in terms of the number of morphemes added to the keyword in a response. In this calculation, the removal of a morpheme from a polymorphemic keyword counts as adding a morpheme. Thus, the morpheme complexity of *vis-(u)al-ly* based on *vision* is 3. Inflectional and non-modifying morphemes are included in the calculation. Table 4 tabulates the mean complexity of the responses.

Table 4: Average complexity of the responses in terms of added morphemes

	By token			By type		
	Means		p	Means		p
DNK vs ISR	1.210	1.270	0.289	1.481	1.501	0.790
DNK vs SLO	1.210	1.246	0.483	1.481	1.427	0.420
ISR vs SLO	1.270	1.246	0.730	1.501	1.427	0.318

As can be seen, the informant groups are fairly close to one another, and the differences are not significant statistically. The fact that the per-token means are lower than the corresponding per-type means shows that the majority of responses contained only one added morpheme.

4.2 *Analysis of modification types*

Table 5 summarizes the findings for modification in terms of item per keyword per informant. As in Table 2, only the informants who actually gave a response to a keyword have been taken into account. It must be noted that the modification types do not exclude each other in that the informants might use all three methods in one single response. Moreover, one single response may well contain several instances of the same modification type except – quite naturally – when the keyword is used as a premodifier. None of the keywords used as premodifier was repeated within the same response, and such a repetition would unlikely make sense. Inflectional morphemes are excluded from the calculation. Inflection is dealt with separately in §4.4 below.

Table 5: Average frequency of the modification types

	By token				By type			
	keyword premo-dified	Key-word post-modi-fied	keyword used as premo-difier	non-modifi-cational response	keyword premo-dified	keyword postmo-dified	keyword used as premo-difier	non-modifica-tional response
DNK	0.938	2.385	0.338	0.002	0.364	0.547	0.112	0.002
ISR	1.141	2.951	0.318	0.046	0.696	1.191	0.172	0.046
SLO	1.013	2.596	0.270	0.000	0.646	1.017	0.114	0.000

It transpires clearly that all informant groups prefer postmodifying the keywords to premodifying them by a ratio of about 3:1 by token and 2:1 by type. Using the keywords as premodifiers is the least preferred method, being by up to a magnitude less frequent than postmodification. This is despite the fact that the keyword *trans* lent itself to being used as a premodifier, and it was the keyword eliciting most responses. Were it not for such an eminently premodifying keyword, the difference between the modification types would likely have been even more marked. Non-modificational responses were given in diminishingly few

instances. Yet they merit special attention and are therefore discussed in a special section below partly because they were totally unexpected and partly because virtually only the Israeli informants provided such responses.

Table 6 shows the statistical significance (p-values) between all three types of modification. No statistics were calculated for the non-modificational responses because of their small amount.

Table 6: Statistical significance of intragroup differences between modification types

	By token			By type		
	keyword premodified vs keyword postmodified	keyword postmodified vs keyword used as premodifier	keyword used as premodifier vs keyword premodified	keyword premodified vs keyword postmodified	keyword postmodified vs keyword used as premodifier	keyword used as premodifier vs keyword premodified
DNK	0.001	0.000	0.066	0.115	0.000	0.014
ISR	0.000	0.000	0.014	0.024	0.000	0.004
SLO	0.001	0.000	0.029	0.152	0.000	0.007

As can be seen, the differences are significant – typically by far – in all but five cases, which are highlighted. Table 7 below enumerates the p-values for the differences between the informant groups with respect to the modification types.

Table 7: Statistical significance of intergroup differences within respect to modification types

	By token			By type		
	keyword premodified	keyword postmodified	keyword used as premodifier	keyword premodified	keyword postmodified	keyword used as premodifier
DNK vs ISR	0.633	0.155	0.875	0.077	0.000	0.296
DNK vs SLO	0.862	0.603	0.601	0.167	0.022	0.962
ISR vs SLO	0.767	0.430	0.684	0.834	0.449	0.295

There is only one case – highlighted – in which the informant groups perform significantly differently from each other: The Danish informants use postmodification by type significantly less than the other two groups. The Danish informants performed most uniformly of the groups by providing relatively few different types of responses. This probably explains why this difference between the informant groups is significant.

4.3 *Compounding*

Since the status of some morphemes is uncertain (see §2.2 above), the calculations concerning compounding presented here are limited to the keywords used as premodifiers. It is hoped that it is uncontroversial that the keywords are roots; thus, when they are used as premodifiers, the resulting constructions should count as compounds beyond reasonable doubt. See §2.2 for the treatment of the keyword *trans*. Table 8 presents the calculations concerning compounding.

Table 8: Frequency of compounding

	By token			**By type**		
	Means		p	Means		p
DNK vs ISR	0.338	0.318	0.875	0.112	0.172	0.296
DNK vs SLO	0.338	0.270	0.601	0.112	0.114	0.962
ISR vs SLO	0.318	0.270	0.684	0.172	0.114	0.295

As can be seen, there is very little difference between the informant groups. The Danish informants do seem to use compounding slightly more than the other two groups token-wise. However, because of their rather uniform responses, also apparent in other measures (cf. e.g. §4.2), they slightly lag behind the other informants' type-wise.

4.4 *Inflection vs derivation*

Table 9 shows the frequency of inflectional morphemes used by the informants. Only explicit inflectional morphemes (plural of nouns (including vowel change as in *man/men*), 3sg of verbs, comparative of adjectives, and superlative of adjectives) have been counted.

Table 9: Frequency of inflection compared to derivational/compounding modification

	By token				**By type**			
	keyword premodi-fied	keyword postmodi-fied	keyword used as premodi-fier	Inflec-tion	keyword premodi-fied	keyword postmodi-fied	keyword used as premodi-fier	Inflec-tion
DNK	0.938	2.385	0.338	0.687	0.364	0.547	0.112	0.121
ISR	1.141	2.951	0.318	0.642	0.696	1.191	0.172	0.190
SLO	1.013	2.596	0.270	0.416	0.646	1.017	0.114	0.190

As can be seen, inflection dwarfs compared to modification. It is type-wise roughly on the same order of magnitude as keywords used as premodifiers, which is the least frequent way of modification used by the informants. Inflection does *outperform* keywords used as

premodifiers token-wise, which suggests that the informants tended to use inflection in the more frequent responses, especially the Danish informants.

4.5 *Non-modificational responses*

At the design stage of this study, only derivation, compounding (including phrasal verbs) and inflection were expected as processes for making new words or new word forms. However, some responses from the Israeli informants and one response from a Danish informant consist of constructions that are not modificational in nature, i.e. they are not derivations, compounds, inflections, or prototypical phrasal verbs. Most of them (all from Israeli informants) are constructions in which *get* serves as an auxiliary or copula verb, or even as a full verb: *get it*, *get older*, *got married*, *got engaged*, *getting jiggy with it*, *got hurt* and *get on one's nerves*. Three responses are noun phrases: *citizen of the world*, based on the keyword *world*; *fear of heights*, based on the keyword *high;* and *on nerve*, based on the keyword *nerve.* Each of these unusual responses was only provided in one instance, though.

5. Conclusion

Below is an overview of the interpretation of the analysis with respect to the individual hypotheses.

Hypothesis 1: There is a statistically significant difference between the informant groups' preferences with respect to word formation.

It must be concluded that the differences between the groups are not statistically significant. Hence it cannot be ascertained that any of the existing differences is caused by the differing mother tongues of the informants (Jarvis 2000). However, this does not necessarily falsify the theory of cross-linguistic influence. The lack of statistically significant differences may be caused by the small sample size of the Israeli and Slovene informants, and may also be caused by the fact that the informants are semi-professional users of English, and thus any differences between the groups that may have existed at an earlier stage of their acquisition of English may have disappeared by now. However, it is remarkable that virtually only the Israeli informants provided phrases or clause-like structures as responses. This begs for explanation and therefore merits further investigation.

Hypothesis 2: The Danish informants use compounding more than the Israeli and Slovene informants.This hypothesis could not be corroborated. There were no significant differences between the informant groups.

Hypothesis 3: All the informant groups use postmodification more than premodification. This hypothesis has been confirmed. Even though the differences between the modification types do not always reach statistically significant levels, the tendencies are clearly in favor of postmodification.

Hypothesis 4: All the informant groups use derivation/compounding more than inflection when producing new "words".

This hypothesis has been confirmed. Unfortunately, there are as yet only remotely sufficient data on English as L2. Since English does not have much inflection left insofar it has many more derivational morphemes than inflectional morphemes, and since there can be only one inflectional morpheme in a word whereas there can be any number of derivatives or compounding elements, it is not surprising that derivation and compounding outweigh

inflection by a large margin. The study will have to be repeated with an L2 that has a richer inflectional system than English to see if inflection might, nevertheless, play a larger role in laymen's word-formation preferences.

The overall conclusion of this study must be that there is more work to be done, for the theory of cross-linguistic influence could be neither refuted nor corroborated on the basis of the available data. Apart from collecting more data from more informants and informant groups (both further L2's and L1's), the design of the study should also be changed so that ambiguities in the responses can be avoided.

References

Abdi, Hervé. 2007. The Bonferonni and Šidák Corrections for Multiple Comparisons. In Salkind, Neil (ed.), *Encyclopedia of Measurement and Statistics*, 103–107. Thousand Oaks, CA: Sage.

American Heritage Dictionary. 2018. (http://www.thefreedictionary.com/) (Accessed 2018-08-31.)

Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.

De Bray, Reginald George Arthur. 1980. *Guide to the South Slavonic languages*. Columbus, Ohio: Slavica Publishers.

Dörnyei, Zoltán. 2014. *Questionnaires in Second Language Research: Construction, Administration, and Processing*, Kindle edition. Abingdon: Taylor and Francis.

Harmon, Mark. 2014. *Practical and Clear Graduate Statistics in Excel: The Excel Statistical Master.* Mark Harmon. Kindle Edition.

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

Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning* 50(2). 245–309.

Jarvis, Scott. 2011. Conceptual transfer: Crosslinguistic effects in categorization and construal. *Bilingualism: Language and Cognition* 14(1). 1–8.

Jarvis, Scott & Pavlenko, Aneta. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.

Johnson, Keith. 2008. *An Introduction to Foreign Language Learning and Teaching*. Upper Saddle River, N.J: Pearson. Kindle Edition.

Lado, Robert. 1957. *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.

Madsen, Richard. 2015. L2 Influence on L1 with Respect to Constituent Order in Translations from English into Danish. In Veselovská, Ludmila & Parrott, Jeffrey K. & Janebová, Markéta (eds.). *Proceedings of the CECIL'S 5: 5th Central european Conference in Linguistics for Postgraduate Students*, 144–150. Olomouc: Palacký University.

Madsen, Richard. Unpublished. Morphological analyzer in Microsoft Excel VBA.

Merriam-Webster. 2018. (https://www.merriam-webster.com/dictionary/combining%20form) (Accessed 2018-08-31.)

Oppenheim, A. N. Bram. 1992. *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Pavlenko, Aneta & Jarvis, Scott. 2002. Bidirectional transfer. *Applied Linguistics* 23. 190–214.

Ramscar, Michael. 2013. Suffixing, prefixing, and the functional order of regularities in meaningful strings. *PSIHOLOGIJA* 46(4). 377–396.

Random House Kernerman Webster's College Dictionary. 2018. (http://www.thefreedictionary.com/) (Accessed 2018-08-31.)

St. Clair, Michelle C. & Monaghan, Padraic & Pamscar, Michael. 2009. Relationships Between Language Structure and Language Learning: The Suffixing Preference and Grammatical Categorization. *Cognitive Science* 33(7). 1317–1329.

Štekauer, Pavol & Valera, Salvador & Körtvélyessy, Lívia. 2012. *Word-Formation in the World's Languages: A Typological Survey*. Cambridge: Cambridge University Press.

Svane, Gunnar Olaf. 1958. *Grammatik der slowenischen Schriftsprache*. Copenhagen: Rosenkilde & Bagger.

Togeby, Ole. 2003. *Fungerer denne sætning? Funktionel dansk sproglære*. [Does this sentence work? A functional grammar of Danish]. Copenhagen: Gads Forlag.

Richard Skultéty Madsen
Section of English Business Communication, Department of Culture and Global Studies
Aalborg University, Denmark
richard@cgs.aau.dk

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/12.pdf. ISSN 1336-782X.