

# Foreigner-directed speech is simpler than native-directed: Evidence from social media

Aleksandrs Berdicevskis

Språkbanken (The Swedish Language Bank), University of Gothenburg

aleksandrs.berdicevskis@gu.se

## Abstract

I test two hypotheses that play an important role in modern sociolinguistics and language evolution studies: first, that non-native production is simpler than native; second, that production addressed to non-native speakers is simpler than that addressed to natives. The second hypothesis is particularly important for theories about contact-induced simplification, since the accommodation to non-natives may explain how the simplification can spread from adult learners to the whole community. To test the hypotheses, I create a very large corpus of native and non-native written speech in four languages (English, French, Italian, Spanish), extracting data from an internet forum where native languages of the participants are known and the structure of the interactions can be inferred. The corpus data yield inconsistent evidence with respect to the first hypothesis, but largely support the second one, suggesting that foreigner-directed speech is indeed simpler than native-directed. Importantly, when testing the first hypothesis, I contrast production of different speakers, which can introduce confounds and is a likely reason for the inconsistencies. When testing the second hypothesis, the comparison is always within the production of the same speaker (but with different addressees), which makes it more reliable.

## 1 Introduction

An important and relatively recent development in sociolinguistics, language evolution and typology is increased interest in sociocognitive determinants of linguistic complexity. Several influential theories (Dahl, 2004; Wray and Grace, 2007; McWhorter, 2007; Trudgill, 2011; Dale and Lupyan, 2012) link the likelihood of a language to maintain or to lose complexity to several social factors, most prominently, the proportion of non-native (L2) speakers in the population and population size. Their main claim can be formulated

as follows: large proportion of L2 speakers and large population size are likely to favour simplification (first of all, morphological simplification), presumably because they inhibit perfect language transmission. While interesting per se, this hypothesis can also be viewed as part of a larger question: to what extent do languages adapt to extralinguistic factors (Five Graces Group et al., 2009; Gibson et al., 2019).

There has been accumulated a solid body of evidence that at least partially supports this causal link. The evidence comes from typological studies (Lupyan and Dale, 2010; Sinnemäki, 2009; Bentz and Winter, 2013; Bentz et al., 2015; Szmeccsanyi and Kortmann, 2009), diachronic analyses (Carroll et al., 2012), computational modeling (Reali et al., 2014) and laboratory experiments (Atkinson et al., 2019; Raviv et al., 2019; Berdicevskis and Semenuks, 2020). Nonetheless, the evidence is not entirely consistent. Studies, for instance, yield different results with respect to whether both factors (population size and proportion of L2 speakers) are at play (Sinnemäki and Di Garbo, 2018) or only one of them (Koplenig, 2019), or whether population size does play a role, but solely because it is strongly correlated with the proportion of L2 speakers. Moreover, the alleged mechanism of simplification is not fully clear.

If we focus on the proportion of L2 speakers as the causal factor, then the hypothetical mechanism of its influence can be in very broad strokes represented as follows:

1. L2 speakers, acquiring the language as adults, often learn it imperfectly;
2. imperfect learning leads to simplified linguistic production;
3. simplification by L2 speakers spreads to the whole community.

Of these links, the first is the most accepted one, while the third causes the most doubt. Indeed, even

if L2s do speak a simplified version of a language, how and why does it affect native speakers? And if it does not, we would not expect the simplification to take place (unless the proportion of L2 speakers is extremely large, which is rarely the case).

It has been suggested (Atkinson et al., 2018) that one of the processes that can account for this “missing link” is foreigner-directed speech, which occurs when more proficient (L1) speakers accommodate their production to less proficient (L2) speakers. Atkinson et al. (2018) observe an equivalent of foreigner-directed speech in their artificial language-learning experiment which does lead to community-level language simplification. They argue convincingly that a similar process may occur in real languages, too, but it is an open question whether it actually does. Moreover, while foreigner-directed speech is being actively studied (Uther et al., 2007; Chun et al., 2016; Wiese, 2009; Lev-Ari et al., 2018; Rothermich et al., 2019), I am not aware of any study that would convincingly show, using a corpus of natural production, that foreigner-directed speech is indeed simpler than native-directed.

Coming back to link 2, it has neither been reliably established that production of L2 speakers is on average simpler than that of L1 speakers. In this paper, I attempt to fill in both these gaps using English, French, Italian and Spanish data extracted from a very large internet forum. The resulting corpus has several advantages over most of the existing learner corpora<sup>1</sup>. It represents naturally occurring written production; it is very large both in terms of number of tokens and number of speakers; the native language of every speaker is known; for many messages, it is possible to identify to whom they are addressed.

To sum up, I test two hypotheses:

1. non-native written speech is on average simpler than native;
2. written speech addressed to non-native speakers is on average simpler than that addressed to natives.

Hypothesis 2 can be further nuanced by asking whether the presence and degree of simplification depends on the addresser being an L1 speaker or an L2 speaker. For the theories about socially-facilitated language simplification, the most important sub-hypothesis is whether L1 speakers accom-

<sup>1</sup>The most comprehensive list can be found here: <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>

	ita	fra	spa	eng
L1 words (M)	3.6	6.8	23.1	71.1
L2 words (M)	1.0	3.8	5.7	54.7
L1 speakers (K)	3.1	5.0	17.4	17.4
L2 speakers (K)	1.9	5.9	7.6	39.2

Table 1: The WordReference corpus composition

modate, but I will test all the possibilities.

The operationalizations of *complexity* and *simpler* are discussed in Section 2.2.

## 2 Materials and methods

### 2.1 The WordReference corpus

All the data are extracted from WordReference forums<sup>2</sup>, where users discuss various questions about languages. Importantly, every user has to provide their native language, and this information, alongside with the nickname, is publicly available (and, of course, visible to other users). Furthermore, for English, French, Italian and Spanish there exist forums “English Only”, “Français Seulement”, “Solo Italiano” and “Sólo Español”, where all communication should occur only in the respective language, and this rule is generally observed.

The data were downloaded in March 2019. They were cleaned of hyperlinks, any traces of markup and all symbols that are neither alphanumeric characters used in the given language nor punctuation marks. Users may quote each other’s messages to make it clear what they are responding to, and while sometimes this information can potentially be useful, it is difficult to process: sometimes users respond without quotes, sometimes one message can contain several quotes or nested quotes (A responding to B’s cue which was written in response to C’s cue etc). For this reason, all quotes are also erased. The resulting corpus size (excluding punctuation) is approximately 170M words (for more details see Table 1). Note that in terms of the number of words, L1 production prevails, even in the English subcorpus where the number of L2 speakers is noticeably higher.

The version of the corpus that is analyzed in this paper consists of four tab-separated files (each per language) with nine columns: message id; poster’s nickname; poster’s native language; the message itself; the id of the first message in the thread (if different from message id); topicstarter’s nickname;

<sup>2</sup><https://forum.wordreference.com/>

topicstarter's native language; poster's status (L1 or L2); topicstarter's status. Statuses were inferred from native languages, but since users provide those as free text (and not by selecting a language from a predefined list), some errors are possible. If a user provides several native languages, he or she is considered to be L1 in all of them.

Most threads are started by non-native speakers (which is unsurprising), but the exact proportion of such threads varies: 66% for Italian, 87% for French, 61% for Spanish and 95% for English. Native speakers, however, also ask questions, which may address, for instance, technical terms, theoretical grammar, differences between language varieties (e.g. American vs British English), language variation, correct usage and pronunciation etc. Note that bilingual and multilingual users are also classified as native speakers as long as the forum language is one of the languages listed in their profile. It is possible that they should instead be treated as L2s, or as a separate category, or that differential treatment should be applied, but a thorough manual investigation would be required for an informed decision.

The corpus (openly available) has several important advantages. It contains naturally occurring written production both of L1 and L2 speakers. It is very large (probably the largest of currently existing learner corpora), both with respect to the number of tokens and the number of represented speakers. It contains some information about the interaction structure (which messages are posted in which thread and in which order). In principle, the social ties between forum users can also be reconstructed, cf. (Del Tredici and Fernández, 2018).

That said, the corpus has two important shortcomings. First, the data are noisy, and presumably the L2 production more so. This means that automatic annotation might introduce an unknown bias, especially if it is used for quantitative comparison of L1 and L2 production. Second, while the corpus contains huge amount of messages from many speakers, most messages are very short. These two disadvantages impose limitations on how complexity can be measured.

## 2.2 Measuring complexity

The research questions and the properties of the corpus yield several requirements to how complexity should be operationalized and quantified. With

those requirements in mind, I opt for type-token ratio (TTR; the number of distinct words divided by the total number of words), a simple measure of lexicogrammatical diversity. First, unlike most other measures, it does not require any annotation. Automatic annotation of non-native speech poses several methodological problems (Díaz-Negrillo et al., 2013). Most importantly, it may introduce an unknown bias due to different performances on L1 and L2 (less standard) data. Controlling for this potential bias implies a substantial manual effort and is beyond the scope of this paper. Second, TTR is very responsive (Bentz et al., 2015) and thus well-suited for the analysis of the short texts. Third, TTR correlates quite well both with other, more advanced corpus-based measures and with manually compiled grammar-based measures of complexity (Bentz et al., 2016; Kettunen, 2014). Fourth, it performs no worse than average in a recent comparison of several corpus-based complexity measures (Berdicevskis et al., 2018).

The main drawback of TTR is that it is very sensitive to text size. To control for that, I use it to compare only the texts of the same size (taking the first  $n$  words of every message). A pilot study suggests that using more advanced TTR-based measures such as HD-D, MTL D (McCarthy and Jarvis, 2010; Koizumi and In'nami, 2012) or moving-average TTR (Covington and McFall, 2010), which are supposed to be more robust, affects the results yielded by plain TTR very little. Since plain TTR is more interpretable, I opt for it.

An alternative approach would be to lump together all messages authored by one speaker, thus creating a slightly larger per-speaker subcorpora, which might make it possible to apply less responsive measures (e.g. word entropy), or even just applying the same measure, TTR, to larger chunks of text. The gain, however, would be small for most speakers, while the downside would be a step away from using a natural unit of analysis (message, a more or less coherent text with certain discursive properties) towards aggregation of data. Instead, I opt for using messages as datapoints and mixed-effect regression (see section 3) for fine-grained analysis without aggregation and with control for non-independence.

Yet another limitation of TTR (as well as most other measures that do not require annotation) is that it cannot distinguish between grammatical complexity (which is the main focus of the theo-

ries discussed in Section 1) and lexical complexity. That is, TTR can decrease because speakers use less different lemmas (lexical complexity), or less different inflected forms (grammatical complexity), or both. However, even with that limitation in mind, the quantitative test of hypotheses 1 and 2 are still valuable, and the individual contributions of lexical and grammatical complexity can be estimated in further studies.

### 3 Comparing complexities

#### 3.1 L1 speakers vs L2 speakers

To test hypothesis 1 (non-native written speech is on average simpler than native), I extract all messages that are at least  $n$  words long and calculate TTR for the first  $n$  words of every such message. I do it for two values of  $n$ : 100 and 200. Smaller values are likely to yield non-robust estimates of TTR, larger values cut off too many messages and make the resulting sample too small. Thus, on the one hand, the 200-word threshold is likely to yield more robust results, but on the other hand, the 100-word thresholds yields more power and can be more informative. The differences between L1 and L2 speakers (for the 100-word threshold) are visualized on Figure 1 using violin plots.

To quantify the observed differences (which are obviously small) and test whether they are significant, I fit a mixed-effect linear regression model with TTR of the message as the dependent variable, speaker's STATUS (L1 vs L2) as main effect (predictor) and the random intercept for speaker id (nickname). The intercept controls for non-independence of the datapoints: several messages can be authored by the same speaker, and there might exist idiosyncratic differences in the complexity of linguistic production between individual speakers, which random effect takes into account. This mixed-effect approach is usually considered superior to methods like usual regression or  $t$ -tests, since it makes it possible to avoid aggregating data (averaging across messages and/or speakers).

I perform the calculations in R 4.0.2 (R Core Team, 2020), using the `lmerTest` package (Kuznetsova et al., 2017) to calculate  $p$ -values and `ggplot2` (Wickham, 2016) to estimate discrimination. In `lme4` (Bates et al., 2015) notation, the model looks as follows:

$$ttr \sim status + (1|speaker)$$

The results of the models for both 200- and 100-word thresholds (which are quite different) are summarized in Table 2.

Obviously, the 100-word threshold yields much more datapoints (though presumably of lower quality). It also yields higher intercept values, which is understandable: as text length increases, TTR decreases, since new types appear at a lower rate than tokens. What is important are different values for slopes (i.e. the effect of STATUS). With the 200-word threshold, the threshold is positive (but not significantly different from zero) for all languages apart from English, where it is significant and negative (i.e. L2 speakers have simpler production, as expected). With the 100-word threshold, the slope is always significant, negative for Italian and positive for all other languages. Note that in all cases the slope is small.

#### 3.2 Foreigner-directed speech

To test hypothesis 2 (non-native directed speech is simpler than native-directed), I extract all messages that are *second* in a thread, i.e. are *first responses* to a posted question (usually the first message in a thread is a question, and other messages are responses to it). I refer to the author of this second message (addresser) as the speaker, and to the author of the first message (addressee) as the topicstarter. The reason for taking only the second messages is that it is only for them we can be fully certain to whom they are addressed. Any other message can potentially be addressed to the topicstarter (answering the original question) or the author(s) of one (or more) messages occurring between the original question and the response, or both, and there is no way to establish that automatically and reliably.

For all selected messages I repeat the procedure described in Section 3.1: if they are long enough, I calculate TTR for the first  $n$  words of every such message,  $n$  being 100 or 200. An additional threshold is that only speakers which has addressed at least one message to an L1 speaker and at least one to an L2 speaker are included. The differences between the complexity of speech addressed to L1 and L2 speakers (for the 100-word threshold) are visualized on Figures 2 (only when the addressers are L1 speakers) and 3 (only when the addressers are L2 speakers).

This time, we have two factors we are interested in: speaker's status and topicstarter's status (who

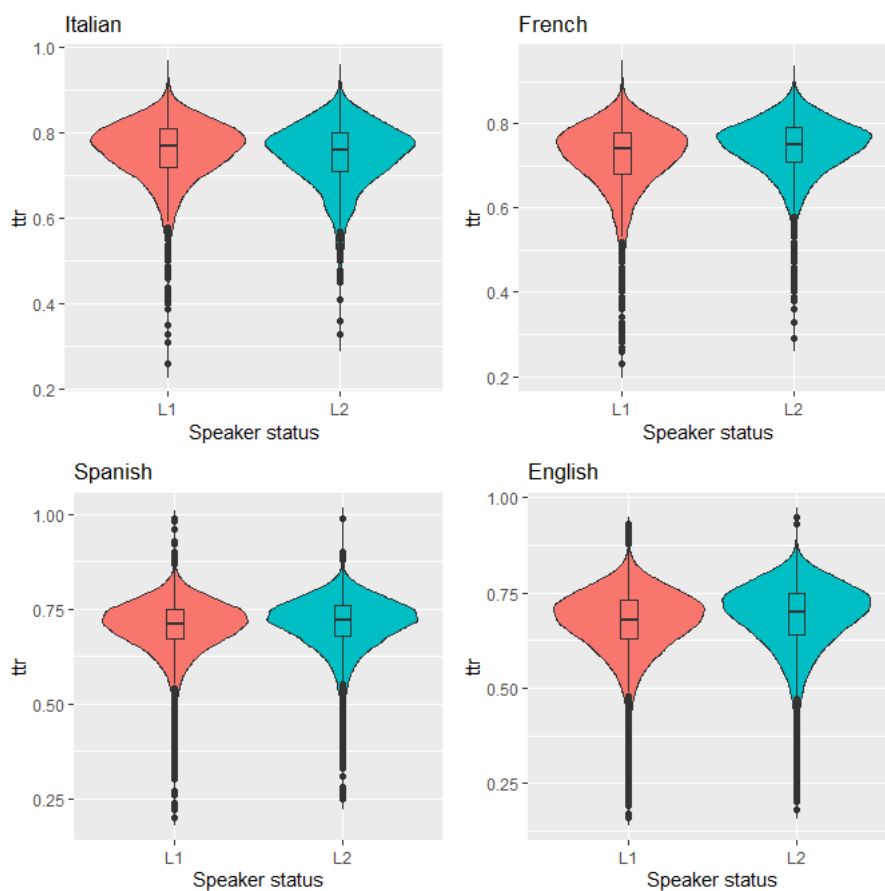


Figure 1: Complexity of L1 and L2 production (raw data, uncontrolled for speaker-specific effects; 100-word threshold)

Language	#messages	#speakers	$\alpha$	$\beta$	SE	$t$ (df)	$p$
Italian-200	2199	451	0.666	-0.005	0.006	-0.8 (358)	0.450
French-200	2742	573	0.628	0.009	0.005	1.9 (439)	0.064
Spanish-200	10675	2033	0.610	0.002	0.003	0.9 (1741)	0.350
English-200	29006	4561	0.585	-0.007	0.002	-3.4 (3461)	<0.001*
Italian-100	11033	1426	0.755	-0.008	0.003	-2.9 (1046)	0.004*
French-100	20787	2539	0.723	0.011	0.002	5.1 (1603)	<0.001*
Spanish-100	61001	6944	0.706	0.004	0.001	3.2 (4415)	0.002*
English-100	258016	18646	0.683	0.003	0.001	-3.3 (14930)	<0.001*

Table 2: Summary of the mixed-effects model for the two thresholds: TTR as predicted by STATUS (L1 vs L2) with random intercept for speaker.  $\alpha$  = intercept,  $\beta$  = slope, all other columns pertain to slope. Asterisks denote significance at the 0.05 level.

Predictor	200-word threshold				100-word threshold			
	Coef	SE	<i>t</i> (df)	<i>p</i>	Coef	SE	<i>t</i> (df)	<i>p</i>
Italian	messages: 98, speakers: 61				messages: 686, speakers: 245			
(Intercept)	0.680	0.012	57 (82)	<0.001*	0.753	0.005	145 (380)	<0.001*
SP:L2	-0.016	0.032	-0.5 (85)	0.620	0.009	0.021	0 (336)	0.662
TS:L2	-0.030	0.015	-2 (88)	0.046*	-0.024	0.006	-4 (677)	<0.001*
SP:L2 x TS:L2	-0.021	0.047	-0.443 (92)	0.659	-0.011	0.022	0 (680)	0.623
French	messages: 166, speakers: 88				messages: 2130, speakers: 410			
(Intercept)	0.601	0.021	59 (162)	<0.001*	0.718	0.005	131 (1349)	<0.001*
SP:L2	0.073	0.040	2 (151)	0.072	0.031	0.018	2 (1420)	0.084
TS:L2	0.003	0.021	0 (157)	0.898	-0.013	0.005	-2 (2097)	0.016*
SP:L2 x TS:L2	-0.075	0.052	-1 (161)	0.154	-0.013	0.019	-1 (1954)	0.556
Spanish	messages: 550, speakers: 228				(messages: 4450, speakers: 864)			
(Intercept)	0.606	0.005	133 (223)	<0.001*	0.702	0.002	321 (946)	<0.001*
SP:L2	-0.028	0.022	-1 (421)	0.191	-0.001	0.008	0 (2113)	0.925
TS:L2	-0.016	0.006	-3 (544)	0.006*	-0.013	0.002	-6 (4440)	<0.001*
SP:L2 x TS:L2	0.041	0.026	2 (538)	0.126	-0.003	0.010	0 (4430)	0.773
English	messages: 1629, speakers: 345				messages: 23284, speakers: 1720			
(Intercept)	0.573	0.007	83 (1338)	<0.001*	0.668	0.002	210 (5810)	<0.001*
SP:L2	-0.004	0.017	0 (1317)	0.804	0.017	0.007	2 (11200)	0.016*
TS:L2	-0.012	0.007	-2 (1625)	0.071	-0.007	0.002	-3 (23240)	0.002*
SP:L2 x TS:L2	0.004	0.018	2 (1625)	0.839	-0.020	0.007	-3 (22820)	0.005*

Table 3: Summary of the mixed-effects model for the 100-word threshold: TTR as predicted by STATUS (L1 vs L2) with random intercept for speaker.  $\alpha$  = intercept,  $\beta$  = slope, all other columns pertain to slope. Asterisks denote significance at the 0.05 level.

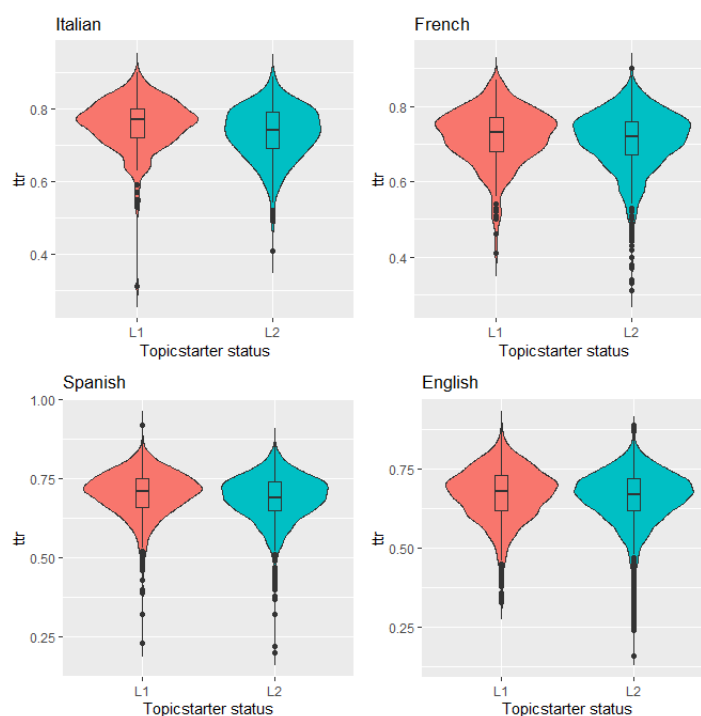


Figure 2: Complexity of L1 production addressed to L1 and L2 speakers (raw data, uncontrolled for speaker-specific effects; 100-word threshold)

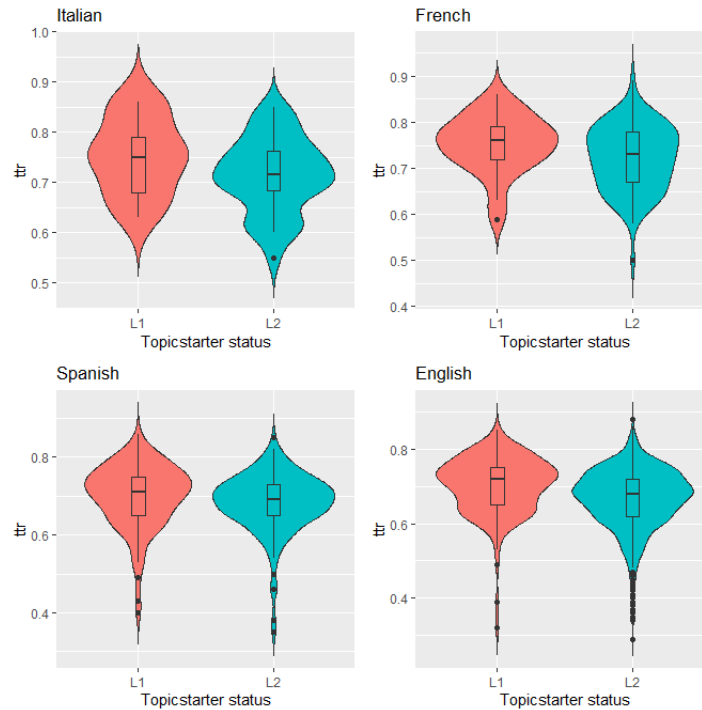


Figure 3: Complexity of L2 production addressed to L1 and L2 speakers (raw data, uncontrolled for speaker-specific effects; 100-word threshold)

speaks to whom). I fit a mixed-effect linear regression model with TTR of the message as the dependent variable, speaker’s status (SP\_STATUS), topicstarter’s status (TS\_STATUS) and their interaction as main effects and the random intercept for speaker id (nickname).

In lme4 notation:

$$ttr \sim sp\_status * ts\_status + (1|speaker)$$

Summaries of the models for both thresholds are provided in Table 3. When it comes to the effect of the topicstarter’s (addressee’s status), the models show much more unity than they did in section 3.1 for speaker’s status. With 200-word threshold, the effect is negative (as expected) for Italian and Spanish and insignificant for French and English. With 100-word threshold, it is always negative and significant. The effect of the speaker’s status is always insignificant apart from English with 100-word threshold, where it is positive (i.e., L2 production is more complex; cf. English-100 in Table 2). The interaction in this case is also significant and negative, implying that L2 speakers accommodate to other L2 speakers more than L1 speakers do (that is not the case for other languages and the other thresholds).

## 4 Discussion

I used the WordReference corpus data to test two hypotheses:

1. non-native written speech is on average simpler than native;
2. written speech addressed to non-native speakers is on average simpler than that addressed to natives.

Generally, the corpus data support hypothesis 2: the results vary across languages and threshold sizes, but the observed effects are either in the direction predicted by hypothesis or insignificant. However, they yield inconsistent results with respect to hypothesis 1. A skeptical reader may ask two questions. Suppose hypothesis 1 is not supported by the data because it is wrong (L2 production is not simpler than L1). Does hypothesis 2 still make sense then? Alternatively, suppose hypothesis 1 is not supported because the chosen method does not really work well on the available data and does not provide reliable complexity estimates. Can the results with respect to hypothesis 2 still be trusted? The answer to both questions is “probably yes”.

The reason why the method described in the paper might work well for the comparison of L1-directed and L2-directed production, but not for the comparison of L1 and L2 production is that in

the latter case what is being compared is always the production of the same speaker (addressed to different persons). In the former case, the production of different speakers is compared, and that might introduce different confounds. It may, for instance, be that some L2 speakers use fewer different forms (TTR decreases), but some make many misspellings (TTR increases)<sup>3</sup>. Other systematic confounds may be present, e.g. that L2 speakers to a larger extent use quotes from textbooks (or some other sources they want to ask a question about) or rely on translation software etc. On the contrary, when the compared texts belong to the same speaker, we can assume that all parameters apart from that of interest (addressee's proficiency) are equivalent, which makes the comparison more reliable.

The same reasons make the comparisons across languages difficult. Cross-linguistic complexity comparisons were not relevant to the theoretical questions of this paper, but are of course still interesting and important. One obvious observation, however, can be made in this respect: English has noticeably lower TTR than the three Romance languages (as expected). In section 2.2, I raised the question whether the observed effects are due to the differences in lexical or grammatical complexity. For the cross-linguistic differences, the reason is likely to be in the domain of the grammar (English inflection is much poorer than Romance; while there is no reason to claim that about English lexicon), which suggests that the other differences are also at least partly due to grammatical complexity.

The answer to the first skeptical question is more hypothetical. If hypothesis 1 is wrong, hypothesis 2 may still be correct. It may be that while L2 speakers are less proficient, their production is not simpler (at least not as measured by TTR): for instance, it can be more variable, thus inflating complexity values. Nonetheless, L1 speakers are sensitive to the differences in proficiency and, being more capable of controlling their production, simplify it in order to accommodate to their interlocutors. That, however, is just a speculation.

The openly-available WordReference corpus can be used for further studies of foreigner-directed speech (and, of course, other phenomena related to non-native acquisition and usage). The understanding of the properties of L2 production can,

<sup>3</sup>One potential source of misspellings are diacritics, especially in French.

for instance, be nuanced by taking into account other relevant factors, such as speakers' mother tongues (Schepens et al., 2020) or social-network structure (Raviv et al., 2020).

## 5 Conclusions

Data from the WordReference corpus (a very large and diverse corpus of naturally occurring written speech in four languages) support the hypothesis that written speech addressed to non-native speakers is on average simpler than that addressed to natives (regardless of who is the addresser: a native or a non-native speaker). There is, however, some variation across languages and measurement thresholds. The same data, however, do not provide conclusive evidence with respect to hypothesis that non-native written speech is on average simpler than native. A probable reason is that the comparison of the production of different speakers is affected by various confounds and not as robust and reliable as comparison within one speaker's production.

The finding that foreigner-directed speech is simpler than native-directed is important for theories of contact-induced language simplification, since it may potentially explain how simplification spreads from non-native speakers to the whole population (including the coming generations).

The scripts that were used to create the corpus, extract the data and perform the statistical analysis, as well as the current version of the corpus are openly available<sup>4</sup>.

## Acknowledgments

I would like to thank the three anonymous reviewers for their comments and suggestions, and my colleagues at the Uppsala university and the University of Gothenburg, as well as the participants of the Evolang-2018 conference for feedback on various earlier versions of this work.

## References

- Mark Atkinson, Gregory J Mills, and Kenny Smith. 2019. [Social group effects on the emergence of communicative conventions and language complexity](#). *Journal of Language Evolution*, 4(1):1–18.
- Mark Atkinson, Kenny Smith, and Simon Kirby. 2018. [Adult learning and language simplification](#). *Cognitive Science*, 42(8):2818–2854.

<sup>4</sup><https://github.com/AleksandrsBerdicevskis/word-reference-corpus>



- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. [Fitting linear mixed-effects models using lme4](#). *Journal of Statistical Software, Articles*, 67(1):1–48.
- Christian Bentz, Tatyana Ruzsics, Alexander Koplenig, and Tanja Samardžić. 2016. [A comparison between morphological complexity measures: Typological data vs. language corpora](#). In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CLALC)*, pages 142–153, Osaka, Japan. The COLING 2016 Organizing Committee.
- Christian Bentz, Annemarie Verkerk, Douwe Kiela, Felix Hill, and Paula Buttery. 2015. [Adaptive communication: Languages with more non-native speakers tend to have fewer word forms](#). *PLOS ONE*, 10(6):1–23.
- Christian Bentz and Bodo Winter. 2013. [Languages with more second language learners tend to lose nominal case](#). *Language Dynamics and Change*, 3(1):1 – 27.
- Aleksandrs Berdicevskis, Çağrı Çöltekin, Katharina Ehret, Kilu von Prince, Daniel Ross, Bill Thompson, Chunxiao Yan, Vera Demberg, Gary Lupyan, Taraka Rama, and Christian Bentz. 2018. [Using universal dependencies in cross-linguistic complexity research](#). In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 8–17, Brussels, Belgium. Association for Computational Linguistics.
- Aleksandrs Berdicevskis and Arturs Semenuks. 2020. Different trajectories of morphological overspecification and irregularity under imperfect language learning. In *The Complexities of Morphology*, pages 283–305. Oxford University Press.
- Ryan Carroll, Ragnar Svare, and Joseph C Salmons. 2012. [Quantifying the evolutionary dynamics of german verbs](#). *Journal of Historical Linguistics*, 2(2):153–172.
- Eunjin Chun, Julia Barrow, and Edith Kaan. 2016. [Native english speakers’ structural alignment mediated by foreign-accented speech](#). *Linguistics Vanguard*, 1(open-issue).
- Michael A. Covington and Joe D. McFall. 2010. [Cutting the gordian knot: The moving-average type–token ratio \(mattr\)](#). *Journal of Quantitative Linguistics*, 17(2):94–100.
- Östen Dahl. 2004. *The growth and maintenance of linguistic complexity*. John Benjamins Publishing.
- Rick Dale and Gary Lupyan. 2012. [Understanding the origins of morphological diversity: The linguistic niche hypothesis](#). *Advances in Complex Systems*, 15(03n04):1150017.
- Marco Del Tredici and Raquel Fernández. 2018. [The road to success: Assessing the fate of linguistic innovations in online communities](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1591–1603, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Ana Díaz-Negrillo, Nicolas Ballier, and Paul Thompson. 2013. *Automatic treatment and analysis of learner corpus data*, volume 59. John Benjamins Publishing Company.
- The Five Graces Group, Clay Beckner, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. 2009. [Language is a complex adaptive system: Position paper](#). *Language Learning*, 59(s1):1–26.
- Edward Gibson, Richard Futrell, Steven P Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. [How efficiency shapes human language](#). *Trends in cognitive sciences*, 23(5):389–407.
- Kimmo Kettunen. 2014. [Can type-token ratio be used to show morphological complexity of languages?](#) *Journal of Quantitative Linguistics*, 21(3):223–245.
- Rie Koizumi and Yo In’nami. 2012. [Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens](#). *System*, 40(4):554 – 564.
- Alexander Koplenig. 2019. [Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers](#). *Royal Society open science*, 6(2):181274.
- Alexandra Kuznetsova, Per B Brockhoff, Rune HB Christensen, et al. 2017. [lmtest package: tests in linear mixed effects models](#). *Journal of statistical software*, 82(13):1–26.
- Shiri Lev-Ari, Emily Ho, and Boaz Keysar. 2018. [The unforeseen consequences of interacting with non-native speakers](#). *Topics in Cognitive Science*, 10(4):835–849.
- Gary Lupyan and Rick Dale. 2010. [Language structure is partly determined by social structure](#). *PLOS ONE*, 5(1):1–10.
- Philip M McCarthy and Scott Jarvis. 2010. [MtlD, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment](#). *Behavior research methods*, 42(2):381–392.
- John McWhorter. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford University Press.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Limor Raviv, Antje Meyer, and Shiri Lev-Ari. 2019. Larger communities create more systematic languages. *Proceedings of the Royal Society B*, 286(1907):20191262.
- Limor Raviv, Antje Meyer, and Shiri Lev-Ari. 2020. The role of social network structure in the emergence of linguistic structure. *Cognitive Science*, 44(8):e12876.
- Florencia Reali, Nick Chater, and Morten Christiansen. 2014. *THE PARADOX OF LINGUISTIC COMPLEXITY AND COMMUNITY SIZE*, pages 270–277. World Scientific.
- Kathrin Rothermich, Havan L. Harris, Kerry Sewell, and Susan C. Bobb. 2019. Listener impressions of foreigner-directed speech: A systematic review. *Speech Communication*, 112:22 – 29.
- Job Schepens, Roeland van Hout, and T. Florian Jaeger. 2020. Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, 194:104056.
- Kaius Sinnemäki. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil, and Peter Trudgill, editors, *Language complexity as an evolving variable*, pages 126–140. Oxford University Press.
- Kaius Sinnemäki and Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology*, 9:1141.
- Benedikt Szmrecsanyi and Bernd Kortmann. 2009. The morphosyntax of varieties of english worldwide: A quantitative perspective. *Lingua*, 119(11):1643 – 1663. The Forests behind the Trees.
- Peter Trudgill. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press.
- Maria Uther, Monja Knoll, and Denis Burnham. 2007. Do you speak e-ng-l-i-sh? a comparison of foreigner- and infant-directed speech. *Speech Communication*, 49(1):2 – 7.
- Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Heike Wiese. 2009. Grammatical innovation in multi-ethnic urban europe: New linguistic practices among adolescents. *Lingua*, 119(5):782 – 806.
- Alison Wray and George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3):543 – 578. The Evolution of Language.