Uncovering Cross-linguistic Morphosyntactic Transfer in Second Language Learning

Many learners find that the morphosyntax of their first language (L1) affects how they learn a second (L2). The reasons, mechanisms, and extent of such influence are debated. Prior studies indeed show complex patterns of both positive (facilitatory) and negative (inhibitory) transfer, varying by morphosyntactic phenomenon and language pair tested (MacWhinney, 1992). Unfortunately, such studies tend to focus on narrowly-defined phenomena in a handful of language pairs (cf. Mitchell et al., 2019), making it difficult to draw a clear picture regarding *whether and which morphosyntactic features are consistently transferred independent of L1-L2 pair*. The few larger-scale experiments mostly focus on character/lexical co-occurrence specific to the language pairs of interest (Malmasi & Dras, 2014).

This study addresses these limitations by using machine learning to identify consistent patterns of transfer (if any) in L2 morphosyntax across 275 L1-L2 pairs. Transfer effects should allow one to distinguish L2 morphosyntax produced by speakers with different L1 backgrounds (Malmasi et al., 2017). Possible outcomes include: no linguistic feature consistently transfers (positively or negatively) across L1-L2 pairs; all linguistic features transfer in many/most cases; or only some aspects of morphosyntax transfer, such as early-learned or UG-implicated phenomena (Mitchell et al., 2019). We note that our focus on common patterns across languages means that we will necessarily miss unusual interactions between specific L1-L2 pairs, which are better assessed in more targeted studies. Additionally, our method assesses transfer holistically and does not cleanly distinguish positive from negative transfer.

Data consisted of 117,163 essays written in L2s, compiled from published learner corpora. While demographics about the writers are minimal – limiting some kinds of analyses – the data set is rich in L1s and L2s (Fig. 1). To determine whether we can detect consistent morphosyntactic transfer across L1-L2 language pairs, we used an automatic parser to identify part-of-speech (POS) tags and dependency relations, which are largely theory-neutral and can be automatically annotated (Berzak et al., 2014; De Marneffe et al., 2021; van der Goot et al., 2021). We then trained a ridge classifier to identify L1s from the patterns of 3-grams of POS tags and 3-grams of dependency relations in the combined data set. The model achieved much better performance than several different baselines (Table 1).

To study what morphosyntactic features are transferred from L1 to L2, we designed a rich set of hand-curated features at the raw text (e.g., the number of words), morphological (e.g., aspect of the verb), and syntactic (e.g., distribution of subordinate clauses) levels. These features were used for L1 classification, where the contribution of each feature was measured as the increase or decrease in classification performance (permutation-adjusted F1 score). The results (Table 2) demonstrated strong transfer effects for certain auxiliary/lexical verb morphological features; in contrast, the relative order of main and subordinate clauses and the average clausal length show little transfer.

We discuss the implications of our results for theories of L2 learning (Mitchell et al., 2019). We also discuss future directions, including potential methods for disentangling positive and negative transfer.

(494 words)

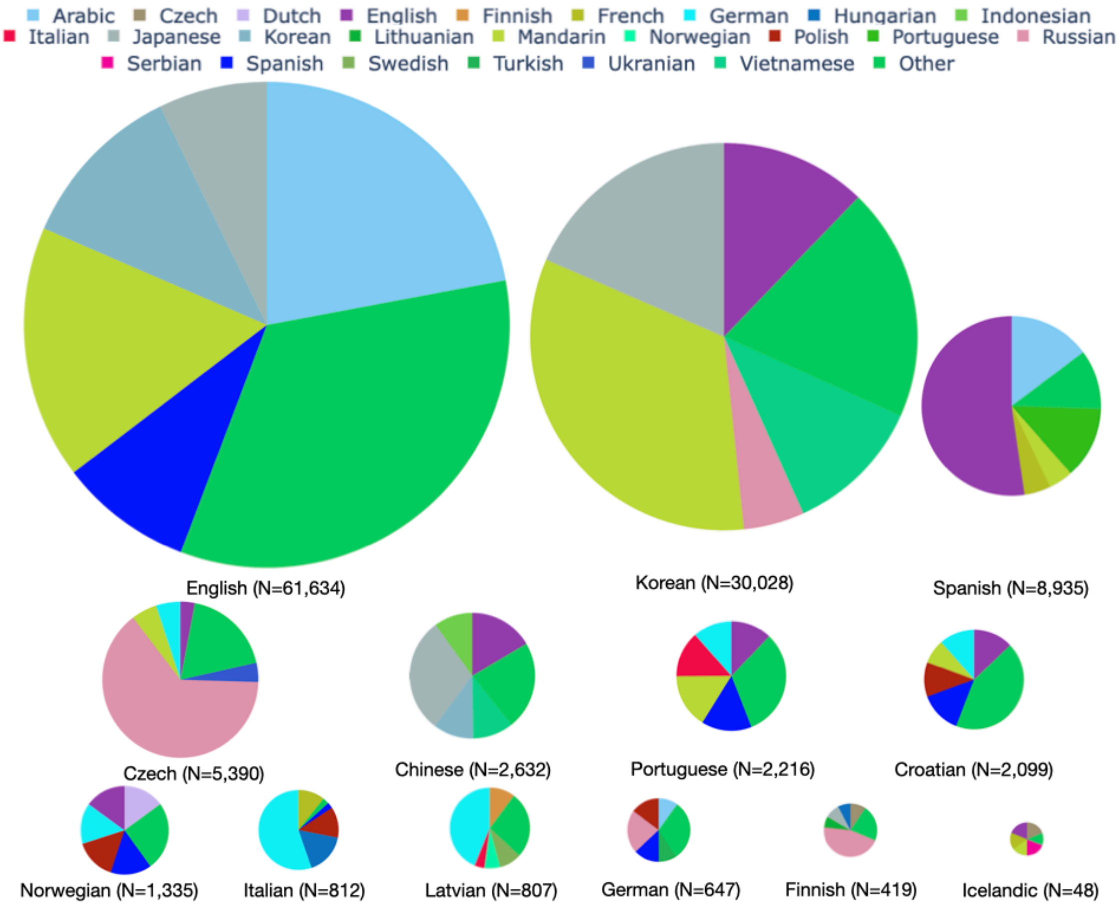

Figure 1: Visualizations of the top 5 most frequent L1 for each of the 13 L2 in our data; pie chart size indicates the size of the data available for the given L2.

Table 1: Classification results using different morphosyntactic representations.

Model	Precision	Recall	F1
Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04
Ridge	**0.41**	**0.41**	**0.41**

Table 2: Examples of most predictive vs. non-predictive features.

Predictive
verb inflection (singular vs plural)
auxiliary inflection (mood; e.g., imperative)
verb inflection (finite vs non-finite)
distribution of dependency relation

Non-predictive
the order of main and subordinate clauses
average clausal length

Berzak, Y., Reichart, R., & Katz, B. (2014). Reconstructing native language typology from foreign language usage. In *Proceedings of the 18th conference on computational natural language learning* (pp. 21–29). **De Marneffe, M. C., Manning, C. D., Nivre, J., & Zeman, D.** (2021). Universal dependencies. *Computational linguistics*, 47(2), 255-308. **MacWhinney, B.** (1992). Transfer and competition in second language learning. In *Advances in psychology* (Vol. 83, pp. 371-390). North-Holland. **Malmasi, S., & Dras, M.** (2014). Language transfer hypotheses with linear SVM weights. In *Proceedings of the 2014 EMNLP* (pp. 1385-1390). **Malmasi, S., Evanini, K., Cahill, A., Tetreault, J., Pugh, R., ... & Qian, Y.** (2017). A report on the 2017 native language identification shared task. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 62-75). **Mitchell, R., Myles, F., & Marsden, E.** (2019). *Second language learning theories.* Routledge. **van der Goot, R., Üstün, A., Ramponi, A., Sharaf, I., & Plank, B**. (2021). Massive Choice, Ample Tasks (MaChAmp): A Toolkit for Multi-task Learning in NLP. In *Proceedings of the 16th Conference of the European Chapter of the ACL: System Demonstrations* (pp. 176-197).