PASTRIE: A Corpus of Prepositions Annotated with Supsersense Tags in Reddit International English

Michael Kranzlein **Emma Manning** **Siyao Peng**
Shira Wein **Aryaman Arora** **Nathan Schneider**
Georgetown University
{mmk119, esm76, sp1184, sw1158, aa2190, nathan.schneider}@georgetown.edu

Abstract

We present the Prepositions Annotated with Supsersense Tags in Reddit International English ("PASTRIE") corpus, a new dataset containing manually annotated preposition supersenses of English data from presumed speakers of four L1s: English, French, German, and Spanish. The annotations are comprehensive, covering all preposition types and tokens in the sample. Along with the corpus, we provide analysis of distributional patterns across the included L1s and a discussion of the influence of L1s on L2 preposition choice.

1 Introduction

It is well-established that one's native language ("L1") leaves traces in second language ("L2") word choice and grammar, including subtle aspects of the use of function words such as prepositions—even for highly proficient L2 speakers (Lowie and Verspoor, 2004; Kujalowicz, 2005; Mueller, 2011; Nacey and Graedler, 2015). However, past corpus studies of L2 writing have had no way to control for the *meaning* of these grammatical items in context on a large scale. In this work, we describe a new corpus, PASTRIE,[1] consisting of English Reddit posts and comments (collectively "documents") that have been manually annotated with preposition supersenses.[2] Following Schneider et al. (2018), the annotations also cover possessives, prepositional multiword expressions ("MWEs"), and infinitives.[3] Examples of annotated sentences appear in (1) and (2).

(1) I was just **on**/Locus it **to**/Purpose find the Copenhagen deal and couldn't find it **at_first**/Time .

(2) Right **at**/Time the moment when that geyser **of**/Stuff light erupts **from**/Source the edge **of**/Whole the screen , we hear a massive rumble come from/Source the door , which was **in**/Locus that direction .

The annotations are comprehensive, covering all types and tokens of prepositional expressions, totaling 2400 tokens out of the 22.5k token corpus. The documents are drawn from the larger Reddit-L2 corpus (Rabinovich et al., 2018), which consists of English Reddit data of speakers of many different L1s. Our corpus includes English produced by presumed native speakers[4] of English, French, German, and Spanish.

Based on annotators' impressions, the English in the corpus produced by the nonnative speakers is highly fluent and unlike what might be found in learner corpora. This is understandable given that these users are taking it upon themselves to post in an online forum, something early learners are less likely to do. This corpus is not only a new resource for exploring preposition supersenses, but it also addresses an understudied niche of broad-coverage semantics for highly proficient non-native data. Using a large,

[1] PASTRIE is available at https://github.com/nert-nlp/pastrie

[2] Automatic lemmas and part-of-speech tags are also included.

[3] We may sometimes use the word "preposition" loosely to cover all of these categories. When specific analyses are being made, more precise terminology is used.

[4] Information on how L1s were identified is included in §3.1. Following Rabinovich et al. (2018), we simply say "native speakers" or "L1" with the understanding that this is an imperfect assumption.

The 14th Linguistic Annotation Workshop, pages 105–116
Barcelona, Spain (Online), December 12, 2020.

unannotated sample of the Reddit-L2 corpus as well as our semantically-annotated subcorpus, we conduct a preliminary investigation of preposition use among English speakers of different L1 backgrounds, extending Rabinovich et al.'s (2018) analysis of L2 lexical choice. We will release the corpus to facilitate further study of such phenomena.

2 Related Work

2.1 Preposition Supersenses

Supersenses are categories used to place both content and function words into unlexicalized semantic classes (Schneider and Smith, 2015), and have been applied to nouns, verbs, adjectives, and adpositions[5] (Miller, 1990; Fellbaum, 1990; Tsvetkov et al., 2015). Here, we focus on the latter. Though adpositions (which almost always occur as prepositions in English) are considered function words and often treated as less important in natural language processing contexts, Schneider et al. (2018) argue for the semantic value of adpositions and propose the Semantic Network of Adposition and Case Supersenses (SNACS) schema.

SNACS categorizes the use of adpositions and case markers, including English possessives, into 50 coarse-grained supersense classes. Each adposition token is annotated as a contrual construction with two of these supersenses (Hwang et al., 2017). A construal includes a Scene Role and a Function, where the former expresses the adposition's meaning in context and the latter denotes its lexical meaning. An example of construal is shown in (3), a sentence from our corpus.[6] In context, the possessive *my* expresses that the speaker is a member of an organization (the company that employs them), hence a scene role of OrgMember; however, the lexical meaning of a grammatical possessive when not indicating possession expresses a looser relationship between entities, corresponding to the function Gestalt. Scene role and function are drawn from the same inventory of supersenses and are often identical. In the PASTRIE corpus, 72% of annotation targets have the same scene role and function.

(3) This is why **my**/OrgMember⇝Gestalt employer has just finished updating 50 k users **from**/Source XP **to**/Goal Windows 7 .

2.2 Prepositions are Uniquely Challenging for Learners

Prepositions are notoriously difficult for language learners (Takahaski, 1969; Littlemore and Low, 2006; Mueller, 2012), which is one of the motivations for constructing this corpus. In studying English preposition usage patterns of high-proficiency learners with different L1 backgrounds, we aim to learn more about how these speakers' L1s might influence their English preposition usage, and how this information might be used to improve pedagogy. One of the problems with prepositions is that they often seem to convey less meaning than content words such as nouns or verbs, but at the same time can be nuanced and highly polysemous. Erarslan and Hol (2014) observed that "most L1 interference took place in the use of prepositions and vocabulary following it." Nacey and Graedler (2015) found rates of inappropriate preposition choice of 4–5% (out of all prepositions) in two corpora of advanced English speakers with a Norwegian L1. They found learners' oral production as challenging as written production and their analysis of the International Corpus of Learner English ("ICLE"; Granger et al., 2009) provided evidence of speakers' L1s influencing L2 lexical choice in both positive and negative ways.

Mahmoodzadeh (2012) conducted Persian-English translation task-based experiments focused on identifying preposition error types. He found that the intermediate Iranian learners of English made more errors of redundancy or inappropriate use than errors of omission and discussed several transfer-related causes of these errors. Gvarishvili (2013) explored negative L1 interference in English preposition usage and offered advice to language educators for mitigating it, but also suggested that educators take advantage of positive influence by pointing out to students L1 prepositions with similar use as their English counterparts.

[5] Adpositions include prepositions, postpositions, and circumpositions, but since we are concerned with English data, we often only mention "prepositions."

[6] Examples use the notation SceneRole⇝Function. When Scene Role and Function have the same supersense label, we write it only once for conciseness.

The difficulty of acquiring prepositions when learning a new language is also addressed in cognitive studies. Lowie and Verspoor (2004) offered a cognitive discussion of the progression of preposition acquisition in Dutch learners of English; Hung et al. (2018) found a cognitive approach that focuses on both spatial and metaphorical meanings to be effective for teaching English prepositions; and Tyler (2012); Bratož (2014); Wong et al. (2018); Zhao et al. (2020) all advocate for a cognitively driven approach to teaching prepositions as well. Pedagogical approaches for teaching English prepositions are also compared in Mueller (2011, 2012). Given the particular difficulty of preposition acquisition, these cognitive pedagogical approaches and new insights from studying learner data should be put to use to help students.

Automatic grammatical error detection (and correction) is another tool that can aid students and has been widely studied, including for prepositions specifically. Models of native and non-native English have been used for this purpose (De Felice and Pulman, 2008; Tetreault and Chodorow, 2008; Hermet and Alain, 2009; Gamon, 2010), along with parse features (Tetreault et al., 2010) and rule-based features (Chodorow et al., 2007). Graën and Schneider (2017) took advantage of parallel corpora for identifying challenging prepositions for learners; Madnani et al. (2011) proposed a crowdsourcing-based approach for improving evaluation of grammatical error detection systems; and Huang et al. (2016) built a Chinese preposition selection model to aid in identifying errors and correcting them. Making explicit some notion of preposition *meaning*, as we do with the PASTRIE corpus, holds the potential to give more informative corrective feedback.

2.3 Reddit-L2: Our Source Corpus

The Reddit-L2 corpus was published in 2018 alongside an analysis of cognate effects in language produced by non-native speakers of English (Rabinovich et al., 2018). It contains 230M sentences and 3.5B tokens of English data from native and non-native speakers, whose L1s were heuristically identified. It was created by first selecting users with a self-specified country *flair* on a set of subreddits and then gathering additional content from those users on different subreddits. While knowing a user's country does not guarantee that their L1 is the majority language in that country, steps were taken to make this more likely, and the inherent noise in the data is acknowledged in the corpus description.[7] The corpus focuses on large languages (it includes authors with flairs from 31 countries representing the Germanic, Romance, and Balto-Slavic language families) and excludes multilingual countries like Switzerland.

Since being made available, the corpus has primarily been used for native language identification (Goldin et al., 2018; Kumar et al., 2019; Steinbakken, 2019; Sarwar et al., 2020). However, it has also been used in studies of bias in word embeddings and bias against non-native text (Manzini et al., 2019; Zhiltsova et al., 2019), as well as semantic infelicity detection (Rabinovich et al., 2019).

3 Corpus Description

3.1 The PASTRIE Corpus

The PASTRIE corpus consists of 1,155 sentences and 22,484 tokens from 255 Reddit documents sampled from the following languages and countries, with percentages of tokens in parentheses:

- English (24.07%): Australia, New Zealand, UK, US
- French (23.56%): France
- German (28.08%): Austria, Germany
- Spanish (24.29%): Argentina, Mexico, Spain

English was chosen as a baseline for comparisons, and French, German, and Spanish were chosen due to their relative similarity to English and wide availability in the Reddit-L2 corpus. While it's possible that some documents belong to the same Reddit thread, this was not a specific selection criterion.

In the corpus, there are 2,395 annotation targets. Of these targets, 2,193 are single tokens and 202 are prepositional MWEs. Sentence segmentation and tokenization were performed with StanfordNLP (Qi et al., 2018), and annotation targets, including prepositions, possessives, MWEs, and infinitives, were

[7] Further details on the construction of the Reddit-L2 corpus are available in section 3 of Rabinovich et al. (2018).

identified heuristically using the same script used for the STREUSLE corpus (Schneider et al., 2018) and then manually corrected during annotation.

3.2 Annotation

3.2.1 Annotation Process

We organized the annotation effort into smaller samples of data (annotation "tasks") that each included 15 documents, and we annotated a total of 17 tasks. All tasks were assigned documents randomly, and documents of each L1 appeared in each task. Tasks were independently annotated by two different annotators, then adjudicated in a meeting which included both annotators and at least one additional person who led the adjudication. Annotators and adjudicators were not shown the L1s of specific documents.

Four different annotators participated over the course of the project, all of whom were Linguistics graduate students and native English speakers; one additional person, a professor with expertise in the annotation scheme did not annotate, but participated in adjudication meetings, especially in the early stages of the project to ensure accuracy. Over all targets, the two annotators agreed on 59.2% of Scene roles (Cohen's $\kappa = 0.58$) and 68.2% of Function labels ($\kappa = 0.66$). This is lower than the SNACS IAA numbers found in Schneider et al. (2018) (74.4% agreement on Scene, 81.3% on Function); however, those were on a sample from a single text, The Little Prince; our data is likely more difficult due to the wide range of topics and authors on social media, and the use of informal and sometimes non-native language.

After initial annotation and adjudication was complete, we did an additional review to ensure annotations were consistent with version 2.5 of the guidelines (Schneider et al., 2020), since most annotation had been done with previous versions, and to resolve difficult cases that were initially left as open or marked as uncertain.

3.2.2 Challenging Cases

One challenge in annotating the data is that Reddit contains discussion of a wide range of topics, often using jargon that would be understood by members of a given subreddit but was not always familiar to annotators; in these cases, annotators looked up terminology or consulted with others to ensure they understood the sentences. The range of topics also meant that many interesting semantic relationships appeared in the data that had not been seen in the STREUSLE corpus. For example, (4) discusses the details of a video game, where a decision had to be made whether to treat *the game* as personified when annotating *by*.

(4) Only Zin and Gore can be knocked **out_of**/SOURCE **their**/GESTALT charged modes , but those are n't considered **as**/CHARACTERISTIC⇝IDENTITY enraged **by**/EXPERIENCER⇝AGENT the game .

In some cases, adpositions represented ambiguous semantic relationships and adjudicators had to decide on the most likely interpretations. For example, in (5), we considered whether the recipe could be considered a personified ORIGINATOR of the suggestion, in which case the possessive would be annotated ORIGINATOR⇝GESTALT. We decided that while this is a possible interpretation, it was more straightforward to consider the suggestion as part of the recipe, hence WHOLE⇝GESTALT.

(5) Never follow a recipe **'s**/WHOLE⇝GESTALT suggestion **for**/TOPIC how much garlic you should put **in**/GOAL⇝LOCUS .

Finally, as with most social media, Reddit text is largely written in an informal register with little or no editing. While this rarely posed a problem for annotation, there were some cases where it was difficult to discern the intended meaning of a sentence. For example, it is unclear whether (6) is referring to a hypothesis that leads to taking a measurement, or the hypothesis made based on a measurement. We decided that it was most likely either EXPLANATION or PURPOSE, and chose EXPLANATION because it is more general. In (7), the preposition *of* doesn't make sense; we annotated it as a typo for *off*, but it could conceivably be a typo for *on* instead.

(6) You can doubt the hypothesis **for**/EXPLANATION a measurement but you can not doubt the actual measurement .

(7) The easiest is getting a bunch of chickens , a rooster , and live **of**/INSTRUMENT⇝SOURCE eggs .

4 Analysis

4.1 Preposition Usage

The PASTRIE corpus is an annotated subcorpus of a larger initial sample we drew from the Reddit-L2 corpus. This sample of roughly 2,500 documents is a more representative source for analysis of preposition usage and can serve as supplementary data for future annotation.

The statistics of the initial sample are described in table 1 and the statistics of the annotated PASTRIE corpus are described in table 2. We see no alarming deviations in PASTRIE compared to the initial sample. PASTRIE contains more English tokens generated by some L1s than others as a result of the random sampling involved in task generation.

L1	Documents	Tokens	Sentences	Prepositions	Prepositions/Token	Tokens/Doc	Sentences/Doc
English	658	48,529	2,544	5038	10.28%	73.75	3.87
French	677	52,093	2,689	5213	10.01%	76.95	3.97
German	767	69,206	3,681	7380	10.66%	90.23	4.80
Spanish	587	45,488	2,410	4588	10.09%	77.49	4.11

Table 1: Characteristics of the initial sample of the Reddit-L2 corpus which tasks were created from.

L1	Documents	Tokens	Sentences	Prepositions	Prepositions/Token	Tokens/Doc	Sentences/Doc
English	67	5,412	284	579	10.70%	80.78	4.24
French	74	5,297	281	539	10.18%	71.58	3.80
German	74	6,313	334	675	10.69%	85.31	4.51
Spanish	65	5,462	256	602	11.00%	84.03	3.94

Table 2: Characteristics of the PASTRIE corpus, the annotated subset of data.

As shown in table 1, prepositions tend to make up 10–11% of the data. The rate of preposition use is highest for German L1 speakers, followed by English and Spanish, with French being the lowest.[8] While German does have the highest rate of preposition use in the initial sample, the range is only 0.65%. This widens slightly to 0.82% for the annotated portion of the data, which has a slightly different ranking.

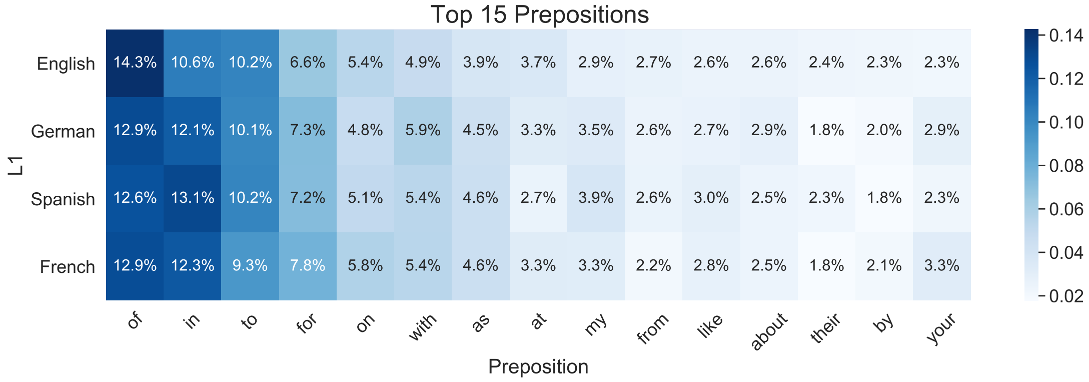


Figure 1: Relative frequencies (among all prepositions for the L1) of 15 most frequent prepositions in the larger unannotated sample of the Reddit-L2 corpus.

Figure 1 indicates that there are some differences in the usage of specific prepositions. Notably, *of* is generated more by native English speakers than by any other speaker, whereas words like *in* and *with* are

[8] The corpus only contains English data. When we mention other languages, we are referring to the L1 of the speaker.

generated more frequently by all other L1 speakers than by native English speakers in this corpus. This may suggest that multiple senses of *of* translate to distinct prepositions in other languages. This could also be due to the mechanics of possession in non-English languages.

Of the top 15 most frequently used prepositions, German L1 speakers collectively had lower relative frequency than at least one other category of L1 speaker for 13/15 prepositions, and a lower relative frequency than at least two other categories of L1 speakers for 11/15 prepositions. Broadly, this suggests that, as shown in Figure 1, German L1 speakers use the 15 most frequently used prepositions less frequently than other L1 speakers, indicating that L1 German learners of English may use a wider variety of prepositions. This could be due to prepositional transfer. The broader use of prepositions by German L1 speakers in this corpus is likely not exclusively due to increased proficiency or near-native English fluency, because German L1 preposition usage does not most closely match the preposition usage of L1 English speakers, as seen in figure 2a. These observations should be taken in context, with caveats of a small sample size and no control for topic and domain of the posts.

4.2 Supersense Usage

The distributions of preposition and supersense usage by L1, depicted in figure 2, are generally comparable in shape. In the preposition usage plot, values are normalized by total frequency of prepositions for each L1. In the supersense plot, values are normalized by the total frequency of the particular construals for each L1.

L1 German Construal Usage Notably, German L1 speakers have the highest number of infrequent *construals*, and the lowest number of infrequent *prepositions*. German also has the longest tail when considering prepositions, while all languages have tails of similar length in the construals plot. The peaked head close to the y-axis indicates a high number of prepositions that are used infrequently and a small number of prepositions that occur frequently. The plot also shows that a few construals and prepositions dominate usage, while most construals and prepositions are infrequently used. The density plot of German L1 construal usage has a less peaked head (/smaller number of low-frequency prepositions) and a less steep decline, meaning German L1 speakers were more likely to use moderate- or high-frequency prepositions.

This supports our claim in Section 4.1, that the range of German L1 preposition usage is being impacted by prepositional transfer. The construal usage by L1 German speakers is not mirroring the construal usage of L1 English speakers, which would be an indication of near-native English fluency and usage, but instead presents differently than the construal usage by all three other L1 speakers.

The density values are normalized by the number of total prepositions generated by each L1, so the less peaked head is not caused by German L1 speakers having generated a larger number of prepositions.

	All		**English**		**French**		**German**		**Spanish**	
Scene Role	Locus	168	`i	45	Topic	35	Topic	49	Locus	51
	Topic	155	Locus	41	Theme	35	Locus	41	Time	39
	Theme	139	Topic	39	Locus	35	CompRef.	38	Theme	38
	`i	137	Gestalt	38	Gestalt	30	Gestalt	37	Goal	36
	Gestalt	127	Theme	37	Circum.	28	Circum.	35	CompRef.	35
Function	Gestalt	325	Gestalt	88	Gestalt	74	Gestalt	87	Gestalt	76
	Locus	242	Locus	55	Locus	55	Locus	60	Locus	72
	Topic	154	Goal	49	Topic	33	Topic	51	Topic	36
	Goal	153	`i	45	CompRef.	30	Goal	44	Time	35
	`i	137	Topic	34	Goal	28	CompRef.	35	Goal	32
Construal	Locus	146	`i	45	Topic	31	Topic	44	Locus	46
	Topic	137	Gestalt	37	Locus	29	Locus	37	Time	35
	`i	137	Locus	34	Gestalt	28	`i	35	Topic	31
	Gestalt	121	Topic	31	`i	28	Gestalt	34	`i	29
	Time	106	Goal	27	`d	22	Circum.	32	Theme	27
	Total	2395	**Total**	579	**Total**	539	**Total**	675	**Total**	602

Table 3: Top scene roles, functions, and construals by L1. In all of the most common construals, the scene role matches the function, so only one supersense is shown.

Most frequent supersenses Table 3 shows that the top labels (for scene role, function, and construal as a whole) across all L1s draw from a limited set of supersenses. The top function supersense across all languages is GESTALT, which is a prototypical function of the English genitives: *of*, *'s*, and the various pronominal forms. Schneider et al. (2018) formulated the SNACS guidelines for these as they were very frequent in past annotated corpora and are highly polysemous; both of these attributes are evident in PASTRIE.

LOCUS is the second most common function in all of the languages. Another example of variation is English's relatively high use of `i, the infinitival uses of *to* and *for*, which are idiomatic to English and thus more difficult to acquire for L2 speakers (Heil and López, 2019).

Supersense distribution comparison Table 4 is a pairwise quantitative comparison of the construals that were encountered between each L1. It shows slight variation in scene role and function and almost no variation in the distribution of construal usage between every pair of languages. This suggests that the general set of meanings that are filled by prepositions is not substantially affected by the L1 of the speaker. Rather, we find that differences manifest in preposition choice for specific construals. An instance of variation in preposition choice for LOCUS is examined below.

L1 vs. **L1**	**Scene**	**Fxn.**	**Cons.**
English vs. German	0.71	0.73	0.61
English vs. French	0.71	0.76	0.61
French vs. Spanish	0.70	0.76	0.59
English vs. Spanish	0.70	0.73	0.61
French vs. German	0.69	0.71	0.60
German vs. Spanish	0.67	0.72	0.61

Table 4: Jaccard similarity coefficients of the multisets of scene roles, functions, and construals between every language pair. Jaccard similarity is a metric of similarity between two sets A, B, defined as $\frac{|A \cap B|}{|A \cup B|}$.

Variation in LOCUS prepositions When the data is examined more narrowly, we do find examples of L1 influence on preposition choice. The most common prepositions used to represent the scene role of LOCUS are *in* and *on*. In the British National Corpus (BNC, 2007), which draws from both formal and informal, written and spoken sources of English, for every instance of *in* there are only 0.35 instances of *on*. In the entire PASTRIE corpus, we find 0.44 instances of *on* for each *in* (disregarding supersense labels).

However, we find that the L1 of non-native speakers significantly skews this ratio when only considering spatial uses of these prepositions. Figure 3 shows that the rate of LOCUS use of *on* relative to *in* is much greater for French (0.91) and German (0.67) than for English (0.23) and Spanish (0.19).

Previous work has observed that spatial relations are categorized differently across languages (Bowerman and Choi, 2001; Feist, 2008) and have complex semantics (Feist, 2000). In English, *in* and *on*, both highly polysemous prepositions, further serve a variety of spatial and metaphoric non-spatial roles (Rice, 1992) which can be difficult for non-native speakers to learn. Language acquisition in regards to motion events between satellite-framed and verb-framed languages is known to be hindered by typological differences (Hickmann and Hendriks, 2010), and more generally due to differences in the semantic fields of spatial markers (Reshöft and Gralla, 2013).

The most likely explanation for these discrepancies across prepositions for LOCUS across L1s is that the semantic fields of spatial markers used in the L1 influences the use of those in the L2. Šeškauskienė and Juknevičienė (2020), from a pedagogical standpoint, do find this effect in Lithuanian L1 speakers' acquisition of English—*in* is learned more readily because it has a clear equivalent in Lithuanian's locative case, while *on* is more difficult because it lacks such an equivalent. Johannes et al. (2016) also examine spatial prepositions in the context of L1 English acquisition, finding that in children the semantic field of *on* is learned much later than that of *in*.

It is possible that the cross-L1 variation in this dataset is due at least in part to varying topics and domains (i.e. which subreddits the documents were sampled from); nevertheless, this example illustrates the utility of SNACS in examining and comparing adposition and case semantics.

5 Conclusion

With data drawn from the Reddit-L2 corpus (Rabinovich et al., 2018), we created PASTRIE, a new corpus of preposition supersense annotations that is publicly available. This corpus adds to existing resources with preposition supersenses and includes annotations of native English data and data produced by L1 speakers of French, German, and Spanish. We demonstrated the applicability of SNACS and the construal analysis to L2 English. We presented detailed discussion of the annotation process, general corpus statistics, and an analysis of usage phenomena across the L1s, including variations between speakers of different L1 backgrounds.

Future work may consider a wider variety of L1s than the typologically similar and closely genetically related languages examined in this work. Computational applications of PASTRIE in natural language understanding (NLU) of non-native English merit further investigation. Finally, corpus-based research such as in this paper can be used to empirically investigate theories of language acquisition.

Acknowledgments

We thank Shuly Wintner, Ella Rabinovich, and Liat Nativ for their assistance in sampling data from the Reddit-L2 corpus. We are grateful to Tripp Maloney, Ryan Mannion, and Sasha Slone for their careful annotation work, and to anonymous reviewers for their feedback. This research was supported in part by NSF award IIS-1812778 and grant 2016375 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

References

BNC. 2007. The British National Corpus, version 3 (BNC XML Edition).

Melissa Bowerman and Soonja Choi. 2001. Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In Melissa Bowerman and Stephen Levinson, editors, *Language Acquisition and Conceptual Development*, number 3 in Language, Culture & Cognition, pages 475–511. Cambridge University Press, Cambridge, UK.

Silva Bratož. 2014. Teaching English locative prepositions: A cognitive perspective. *Linguistica*, 54(1):325–337.

Martin Chodorow, Joel R. Tetreault, and Na-Rae Han. 2007. Detection of grammatical errors involving prepositions. In *Proc. of the Fourth ACL-SIGSEM Workshop on Prepositions*, pages 25–30, Prague, Czech Republic.

Rachele De Felice and Stephen G. Pulman. 2008. A classifier-based approach to preposition and determiner error correction in L2 English. In *Proc. of Coling*, pages 169–176, Manchester, UK.

Ali Erarslan and Devrim Hol. 2014. Language interference on English: Transfer on the vocabulary, tense and preposition use of freshmen Turkish EFL learners. *ELTA Journal*, 2(2):4–22.

Michele I Feist. 2000. *On in and on: An investigation into the linguistic encoding of spatial scenes*. Ph.D. thesis, Northwestern University.

Michele I. Feist. 2008. Space between languages. *Cognitive Science*, 32(7):1177–1199.

Christiane Fellbaum. 1990. English verbs as a semantic net. *International Journal of Lexicography*, 3(4):278–301.

Michael Gamon. 2010. Using mostly native data to correct errors in learners' writing: a meta-classifier approach. In *Proc. of NAACL-HLT*, pages 163–171, Los Angeles, California.

Gili Goldin, Ella Rabinovich, and Shuly Wintner. 2018. Native language identification with user generated content. In *Proc. of EMNLP*, pages 3591–3601, Brussels, Belgium.

Johannes Graën and Gerold Schneider. 2017. Crossing the border twice: reimporting prepositions to alleviate L1-specific transfer errors. In *Proc. of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa, Gothenburg, 22nd May 2017*, pages 18–26, Gothenburg, Sweden.

Sylviane Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. 2009. *International Corpus of Learner English*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Zeinab Gvarishvili. 2013. Interference of L1 prepositional knowledge in acquiring of prepositional usage in English. *Procedia - Social and Behavioral Sciences*, 70:1565–1573.

Jeanne Heil and Luis López. 2019. Acquisition without evidence: English infinitives and poverty of stimulus in adult second language acquisition. *Second Language Research*.

Matthieu Hermet and Désilets Alain. 2009. Using first and second language models to correct preposition errors in second language authoring. In *Proc. of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 64–72, Boulder, Colorado.

Maya Hickmann and Henriëtte Hendriks. 2010. Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics*, 21(2):189–215.

Hen-Hsen Huang, Yen-Chi Shao, and Hsin-Hsi Chen. 2016. Chinese preposition selection for grammatical error diagnosis. In *Proc. of COLING*, pages 888–899, Osaka, Japan.

Bui Phu Hung, Truong Vien, and Nguyen Ngoc Vu. 2018. Applying cognitive linguistics to teaching English prepositions: A quasi-experimental study. *International Journal of Instruction*, 11(3):327–346.

Jena D. Hwang, Archna Bhatia, Na-Rae Han, Tim O'Gorman, Vivek Srikumar, and Nathan Schneider. 2017. Double trouble: the problem of construal in semantic annotation of adpositions. In *Proc. of *SEM*, pages 178–188, Vancouver, Canada.

Kristen Johannes, Colin Wilson, and Barbara Landau. 2016. The importance of lexical verbs in the acquisition of spatial prepositions: The case of in and on. *Cognition*, 157:174–189.

Agnieszka Kujalowicz. 2005. Cross-linguistic influence in the production of German prepositions by Polish learners of English and German. *Studia Anglica Posnaniensia: International Review of English Studies*, 41:187–198.

Sachin Kumar, Shuly Wintner, Noah A. Smith, and Yulia Tsvetkov. 2019. Topics to avoid: demoting latent confounds in text classification. In *Proc. of EMNLP-IJCNLP*, pages 4153–4163, Hong Kong, China.

Jeannette Littlemore and Graham D Low. 2006. *Figurative thinking and foreign language learning*. Springer.

Wander Lowie and Marjolijn Verspoor. 2004. Input versus transfer? - the role of frequency and similarity in the acquisition of L2 prepositions. In Peter Jordens, Michel Achard, and Susanne Niemeier, editors, *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*, 2004 edition, volume 18, pages 77–94. Mouton de Gruyter, Berlin, New York.

Nitin Madnani, Martin Chodorow, Joel Tetreault, and Alla Rozovskaya. 2011. They can help: using crowdsourcing to improve the evaluation of grammatical error detection systems. In *Proc. of ACL-HLT*, pages 508–513, Portland, Oregon, USA.

Masoud Mahmoodzadeh. 2012. A cross-linguistic study of prepositions in Persian and English: The effect of transfer. *Theory and Practice in Language Studies*, 2(4):734–740.

Thomas Manzini, Lim Yao Chong, Alan W Black, and Yulia Tsvetkov. 2019. Black is to criminal as Caucasian is to police: detecting and removing multiclass bias in word embeddings. In *Proc. of NAACL-HLT*, pages 615–621, Minneapolis, Minnesota.

George A. Miller. 1990. Nouns in WordNet: A lexical inheritance system. *International Journal of Lexicography*, 3(4):245–264.

Charles M. Mueller. 2011. English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? *System*, 39(4):480–490.

Charles M Mueller. 2012. *Comparison of an Integrative Inductive Approach, Presentation-and-Practice Approach, and Two Hybrid Approaches to Instruction of English Prepositions*. Ph.D. thesis, University of Maryland.

Susan Nacey and Anne-Line Graedler. 2015. Preposition use in oral and written learner language. *Bergen Language and Linguistics Studies*, 6.

Peng Qi, Timothy Dozat, Yuhao Zhang, and Christopher D. Manning. 2018. Universal Dependency parsing from scratch. In *Proc. of CoNLL*, pages 160–170, Brussels, Belgium.

Ella Rabinovich, Yulia Tsvetkov, and Shuly Wintner. 2018. Native language cognate effects on second language lexical choice. *Transactions of the Association for Computational Linguistics*, 6:329–342.

Ella Rabinovich, Julia Watson, Barend Beekhuizen, and Suzanne Stevenson. 2019. Say anything: automatic semantic infelicity detection in L2 English indefinite pronouns. In *Proc. of CoNLL*, pages 77–86, Hong Kong, China.

Nina Reshöft and Linn Gralla. 2013. On the use of spatial prepositions: Differences in L1 and L2 English. *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use–Proceedings*, 1:389–400.

Sally A Rice. 1992. Polysemy and lexical representation: The case of three English prepositions. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 89–94.

Raheem Sarwar, Attapol T. Rutherford, Saeed-Ul Hassan, Thanawin Rakthanmanon, and Sarana Nutanong. 2020. Native language identification of fluent and advanced non-native writers. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 19(4):1–19.

Nathan Schneider, Jena D. Hwang, Archna Bhatia, Vivek Srikumar, Na-Rae Han, Tim O'Gorman, Sarah R. Moeller, Omri Abend, Adi Shalev, Austin Blodgett, and Jakob Prange. 2020. Adposition and Case Supersenses v2.5: Guidelines for English. *arXiv:1704.02134v6 [cs]*.

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proc. of ACL*, pages 185–196, Melbourne, Australia.

Nathan Schneider and Noah A. Smith. 2015. A corpus and model integrating multiword expressions and supersenses. In *Proc. of NAACL-HLT*, pages 1537–1547, Denver, Colorado.

Inesa Šeškauskienė and Rita Juknevičienė. 2020. Prepositions in L2 written English, or why on poses more difficulties than in. *Nordic Journal of English Studies*, 19(1).

Stian Steinbakken. 2019. *Paying Attention to Native-Language Identification*. Ph.D. thesis, Norwegian University of Science and Technology.

George Takahaski. 1969. Perception of space and the function of certain English prepositions. *Language Learning*, 19(3-4):217–234.

Joel Tetreault, Jennifer Foster, and Martin Chodorow. 2010. Using parse features for preposition selection and error detection. In *Proc. of the ACL 2010 Conference Short Papers*, pages 353–358, Uppsala, Sweden.

Joel R. Tetreault and Martin Chodorow. 2008. The ups and downs of preposition error detection in ESL writing. In *Proc. of Coling*, pages 865–872, Manchester, UK.

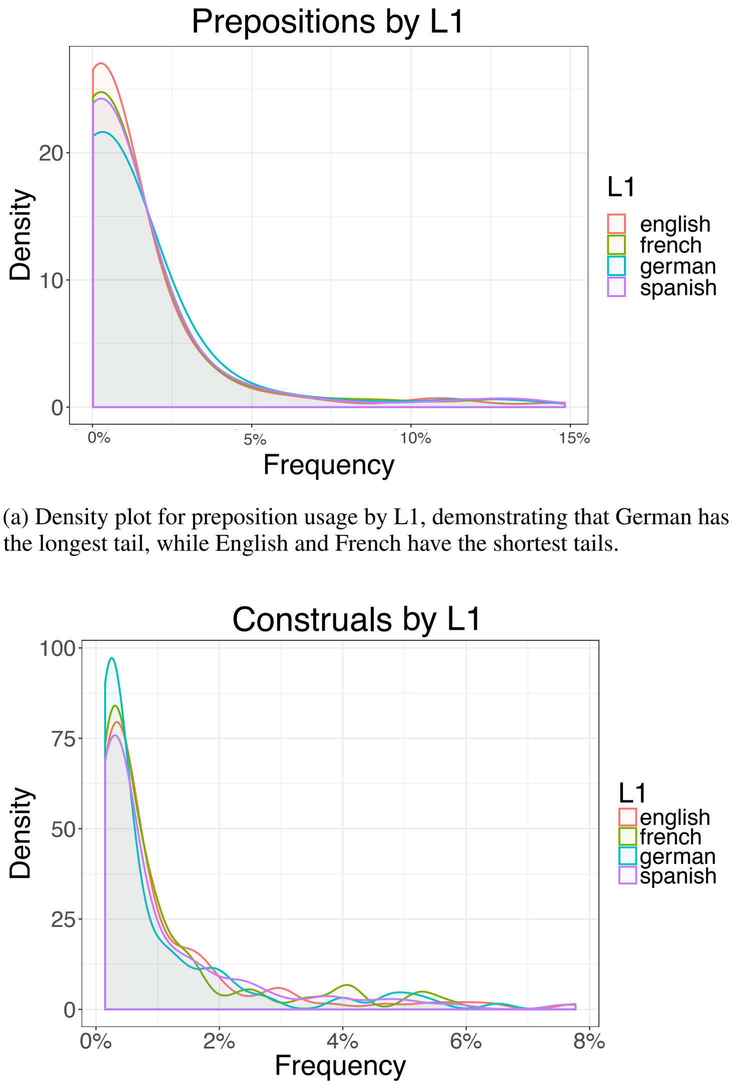
Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer. 2015. Evaluation of word vector representations by subspace alignment. In *Proc. of EMNLP*, pages 2049–2054, Lisbon, Portugal.

Andrea Tyler. 2012. *Cognitive Linguistics and Second Language Learning: Theoretical Basics and Experimental Evidence*. Routledge.

Man Ho Ivy Wong, Helen Zhao, and Brian MacWhinney. 2018. A cognitive linguistics application for second language pedagogy: The English preposition tutor. *Language Learning*, 68(2):438–468.

Helen Zhao, Shuting Huang, Yacong Zhou, and Ruiming Wang. 2020. Schematic diagrams in second language learning of English prepositions - a behavioral and event-related potential study. *Studies in Second Language Acquisition*, pages 1–28.

Alina Zhiltsova, Simon Caton, and Catherine Mulwa. 2019. Mitigation of unintended biases against non-native English texts in sentiment analysis. In *Proceedings for the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, page 12.



(a) Density plot for preposition usage by L1, demonstrating that German has the longest tail, while English and French have the shortest tails.

(b) Density plot of construals by L1. This depicts the number of construals (y-axis) that have a certain frequency (x-axis) in the annotated corpus. Most construals are used rarely, and there are only a few high-frequency construals.

Figure 2: Density plots for prepositions and construals, normalized by total number of prepositions per L1. Recall that a construal is a pair of supersenses—a SCENE ROLE and a FUNCTION.

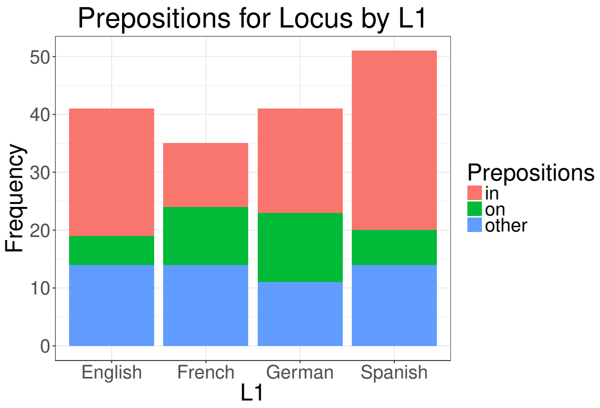


Figure 3: Frequency counts of tokens annotated with LOCUS as the scene role, broken down by native language and preposition type: *in*, *on*, and others.