Norwegian Native Language Identification

Shervin Malmasi◇   **Mark Dras**◇   **Irina Temnikova**♡

◇Macquarie University, Sydney, NSW, Australia
♡Qatar Computing Research Institute, HBKU, Qatar

`shervin.malmasi@mq.edu.au, mark.dras@mq.edu.au`
`itemnikova@qf.org.qa`

Abstract

We present a study of Native Language Identification (NLI) using data from learners of Norwegian, a language not yet used for this task. NLI is the task of predicting a writer's first language using only their writings in a learned language. We find that three feature types, function words, part-of-speech n-grams and a hybrid part-of-speech/function word mixture n-gram model are useful here. Our system achieves an accuracy of 79% against a baseline of 13% for predicting an author's L1. The same features can distinguish non-native writing with 99% accuracy. We also find that part-of-speech n-gram performance on this data deviates from previous NLI results, possibly due to the use of manually post-corrected tags.

1 Introduction

Native Language Identification (NLI) is the task of identifying a writer's native language (L1) based only on their writings in a second language (the L2). NLI works by identifying language use patterns that are common to groups of speakers of the same native language. This process is underpinned by the presupposition that an author's L1 disposes them towards certain language production patterns in their L2, as influenced by their mother tongue. This relates to cross-linguistic influence (CLI), a key topic in the field of Second Language Acquisition (SLA) that analyzes transfer effects from the L1 on later learned languages (Ortega, 2009).

It has been noted in the linguistics literature since the 1950s that speakers of particular languages have characteristic production patterns when writing in a second language. This language transfer phenomenon has been investigated independently in various fields from different perspectives, including qualitative research in SLA and more recently though predictive computational models in NLP (Jarvis and Crossley, 2012).

Recently this has motivated studies in Native Language Identification (NLI), a subtype of text classification where the goal is to determine the native language (L1) of an author using texts they have written in a second language or L2 (Tetreault et al., 2013).

The motivations for NLI are manifold. The use of such techniques can help SLA researchers identify important L1-specific learning and teaching issues. In turn, the identification of such issues can enable researchers to develop pedagogical material that takes into consideration a learner's L1 and addresses them. It can also be applied in a forensic context, for example, to glean information about the discriminant L1 cues in an anonymous text. In fact, recent NLI research such as that related to the work presented by Perkins (2014) has already attracted interest and funding from intelligence agencies (Perkins, 2014, p. 17).

While most NLI research to date has focused on English L2 data, there is a growing trend to apply the techniques to other languages in order to assess their cross-language applicability (Malmasi and Dras, 2014c).

The current work presents the first NLI experiments on Norwegian data using a corpus of examination essays collected from learners of Norwegian, as described in section 3. Given the differences between English and Norwegian (which we outline in section 2.1), the main objective of the present study is to determine if NLI techniques previously applied to L2 English can be effective for detecting L1 transfer effects in L2 Norwegian.

Another unique aspect of this data is the availability of manually corrected part-of-speech (POS) tag annotations. This is something that has not been generally considered in previous NLI research and we aim to analyze how these results compare to previous studies in this regard.

2 Background and Related Work

NLI work has been growing in recent years, using a wide range of syntactic and more recently, lexical features to distinguish the L1. A detailed review of NLI methods is omitted here for reasons of space, but a thorough exposition is presented in the report from the very first NLI Shared Task that was held in 2013 (Tetreault et al., 2013).

Most English NLI work has been done using two corpora. The *International Corpus of Learner English* (Granger et al., 2009) was widely used until recently, despite its shortcomings[1] being widely noted (Brooke and Hirst, 2012). More recently, TOEFL11, the first corpus designed for NLI was released (Blanchard et al., 2013). While it is the largest NLI dataset available, it only contains argumentative essays, limiting analyses to this genre.

Research has also expanded to use non-English learner corpora (Malmasi and Dras, 2014a; Malmasi and Dras, 2014c). Recently, Malmasi and Dras (2014b) introduced the Jinan Chinese Learner Corpus (Wang et al., 2015) for NLI and their results indicate that feature performance may be similar across corpora and even L1-L2 pairs. In this work we attempt to follow this exploratory pattern by extending NLI research to Norwegian, which has not yet been studied for this task.

NLI is now also moving towards using linguistic features to generate SLA hypotheses. Swanson and Charniak (2014) approach this by using both L1 and L2 data to identify features exhibiting non-uniform usage in both datasets, creating lists of candidate transfer features. Malmasi and Dras (2014d) propose a different method, using linear SVM weights to extract lists of overused and underused linguistic features for each L1 group.

Many of these studies have investigated using syntactic information such as parse trees or part-of-speech (POS) tags as classification features (Kochmar, 2011). This is generally achieved by using taggers and parsers based on statistical models to automatically annotate the documents. For example, Tetreault et al. (2012) use the Stanford Tagger (Toutanova et al., 2003) to extract POS tags from the TOEFL11 data.

One issue to consider here is that the models used by these statistical taggers are trained on well-formed text from a standard variety of the language written by native speakers (*e.g.* news articles). When tested on such data, the models generally achieve high accuracies of 95% or higher. However, it cannot be assumed that these tools will achieve similar levels of accuracy on learner data, a distinct genre which they were not trained on.

This is a consideration that has not gone unnoticed and several researchers have investigated this question. Van Rooy and Schäfer (2002) investigated this issue and report that "learner spelling errors contributed substantially to tagging errors", causing up to 38% of the tagging errors. Dıaz-Negrillo et al. (2010) argue that the properties of learner language are systematically different from those assumed for the standard variety of the language and that this interlanguage cannot be considered a noisy variant of the native language. Instead of viewing this as a robustness issue, they suggest that a new POS model for learner language may be more suitable. Based on the results of their empirical analysis they highlight several issues with standard POS models and they propose a new tripartite POS annotated model that encodes properties based on the lexical stem, distribution and morphology.

This evidence points to a performance degradation on learner data and suggests that the POS annotations used in many previous studies are vulnerable to tagging errors. Such errors could reduce their efficacy in distinguishing the different syntactic patterns used by different L1 groups. The availability of post-corrected POS tags in our data, as described in §3, can provide some insight into how much this issue affects NLI by comparing its performance with previously reported results.

2.1 Norwegian

Norwegian can be considered as one of the mainland Scandinavian languages. Along with Danish and Swedish, these languages share their heritage and have descended from a common Nordic language. Even today, a degree of mutual intelligibility continues to exist among these languages.

Norwegian itself is written in two distinguishable forms: Bokmål and Nynorsk with the former being more commonly used for writing, including in our data. The language has a number of properties that make it interesting to examine for NLI.

Norwegian grammar shares many similarities with English since both are Germanic languages. However, a number of differences also exist.

Norwegian has three genders: male, female and neuter. Definite and indefinite articles also ex-

---

[1] The issues exist as the corpus was not designed for NLI.

ist for all three genders, but the definite article is added to nouns as a suffix. Nouns are categorized by gender and in addition to definiteness, they are also inflected for plurality. Pronouns are classified by gender, person and number; they are also declined in nominative or accusative case. Adjectives must agree with gender of their head nouns and are also marked for plurality and definiteness. Norwegian verbs, although not marked for person or plurality, can have several different tenses and moods, leading to a rich morphology.

An important point to consider here is that this additional complexity also increases the possibility and number of potential learner errors. A more in-depth exposition of Norwegian syntax and morphology can be found in Haugen (2009).

3 Data

In this study we use data from the ASK Corpus (*Andrespråkskorpus*, Second Language Corpus). The ASK Corpus (Tenfjord et al., 2013; Tenfjord et al., 2006b; Tenfjord et al., 2006a) is a learner corpus composed of the writings of learners of Norwegian. These texts are essays written as part of a test of Norwegian as a second language. Each text also includes additional metadata about the author such as age or native language. An advantage of this corpus is that all the texts have been collected under the same conditions and time limits. The corpus also contains a control subcorpus of texts written by native Norwegians under the same test conditions. The corpus also includes error codes and corrections, although we do not make use of this information here.

There are a total of 1,700 essays written by learners of Norwegian as a second language with ten different first languages: German, Dutch, English, Spanish, Russian, Polish, Bosnian-Croatian-Serbian, Albanian, Vietnamese and Somali. The essays are written on a number of different topics, but these topics are not balanced across the L1s.

Detailed word level annotations (lemma, POS tag and grammatical function) have been first obtained automatically using the Oslo-Bergen tagger. These annotations have then been manually post-edited by human annotators since the tagger's performance can be substantially degraded due to orthographic, syntactic and morphological learner errors. These manual corrections can deal with issues such as unknown vocabulary or wrongly disambiguated words.
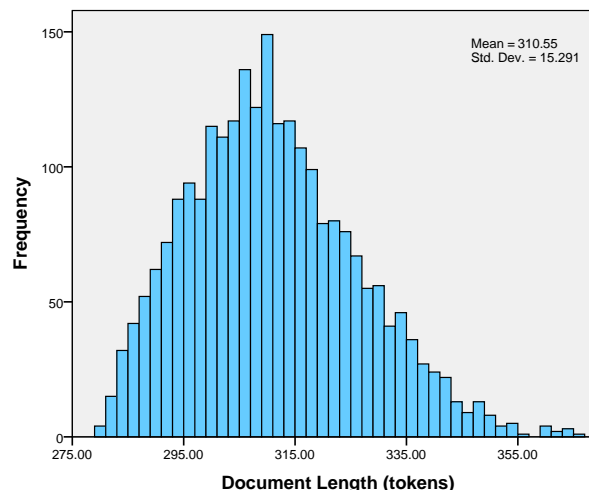

Figure 1: A histogram of the number of tokens per document in the dataset that we generated.

In this work we extracted 750k tokens of text from the ASK corpus in the form of individual sentences. Following the methodology of Brooke and Hirst (2011) and Malmasi and Dras (2014b), we randomly select and combine the sentences from the same L1 to generate texts of approximately 300 tokens on average, creating a set of documents suitable for NLI. This methodology ensures that the texts for each L1 are a mix of different authorship styles, topics and proficiencies. It also means that all documents are similar and comparable in length.

The 10 native languages and the number of texts generated per class are listed in Table 1. In addition to these we also generate 250 control texts written by natives. A histogram of the number of tokens per document is shown in Figure 1. The documents have an average length of 311 tokens with a standard deviation of 15 tokens.

3.1 Part-of-Speech Tagset

The ASK corpus uses the Oslo-Bergen tagset[2] which has been developed based on the Norwegian Reference Grammar (Faarlund et al., 1997).

Here each POS tag is composed of a set of constituent morphosyntactic tags. For example, the tag `subst-appell-mask-ub-fl` signifies that the token has the categories "noun common masculine indefinite plural". Similarly, the tags `verb-imp` and `verb-pres` refer to imperative and present tense verbs, respectively.

---

[2]`http://tekstlab.uio.no/obt-ny/english/tagset.html`

Native Language	Documents
Albanian	121
Dutch	254
English	273
German	280
Polish	281
Russian	257
Serbian	259
Somali	90
Spanish	243
Vietnamese	100
Total	2,158

Table 1: The 10 L1 classes included in this experiment and the number of texts we generated for each class.

Given its many morphosyntactic markers and detailed categories, the ASK dataset has a rich tagset with over 300 unique tags.

4 Experimental Methodology

In this study we employ a supervised multi-class classification approach. The learner texts are organized into classes according to the author's L1 and these documents are used for training and testing in our experiments. A diagram conceptualizing our NLI system is shown in Figure 2.

4.1 Classifier

We use a linear Support Vector Machine to perform multi-class classification in our experiments. In particular, we use the LIBLINEAR[3] package (Fan et al., 2008) which has been shown to be efficient for text classification problems such as this. More specifically, it has been demonstrated to be the most effective classifier for this task in the 2013 NLI Shared Task (Tetreault et al., 2013).

4.2 Evaluation

In the same manner as many previous NLI studies and also the NLI 2013 shared task, we report our results as classification accuracy under k-fold cross-validation, with $k = 10$. In recent years this has become a *de facto* standard for reporting NLI results.

5 L1 Identification Experiment

We experiment using three syntactic feature types described in this section. As the ASK corpus is not balanced for topic, we do not consider the use of lexical features such as word n-grams in this study. Topic bias can occur as a result of the subject matters or topics of the texts to be classified not evenly distributed across the classes (Koppel et al., 2009). For example, if in our training data all the texts written by English L1 speakers are on topic A, while all the French L1 authors write about topic B, then we have implicitly trained our classifier on the topics as well. In this case the classifier learns to distinguish our target variable through another confounding variable.

Norwegian Function Words As opposed to content words, function words are topic-independent grammatical words that indicate the relations between other words. They include determiners, conjunctions and auxiliary verbs. Distributions of English function words have been found to be useful in studies of authorship attribution and NLI. Unlike POS tags, this model analyzes the author's specific word choices.

In this work we used a list of 176 function words obtained from the distribution of the Apache Lucene search engine software.[4] This list includes stop words for the Bokmål variant of the language and contains entries such as *hvis* (whose), *ikke* (not), *jeg* (I), *så* (so) and *hjå* (at). We also make this list available on our website.[5]

In addition to single function words, we also extract function word bigrams, as described by Malmasi et al. (2013). Function word bigrams are a type of word n-gram where content words are skipped: they are thus a specific subtype of skip-gram discussed by Guthrie et al. (2006). For example, the sentence *We should all start taking the bus* would be reduced to *we should all the*, from which we would extract the n-grams.

Part-of-Speech n-grams In this model POS n-grams of order 1–3 were extracted. These n-grams capture small and very local syntactic patterns of language production and were used as classification features. Previous work and our experiments showed that sequences of size 4 or greater achieve

---

[3]http://www.csie.ntu.edu.tw/%7Ecjlin/liblinear/

[4]https://github.com/apache/lucene-solr
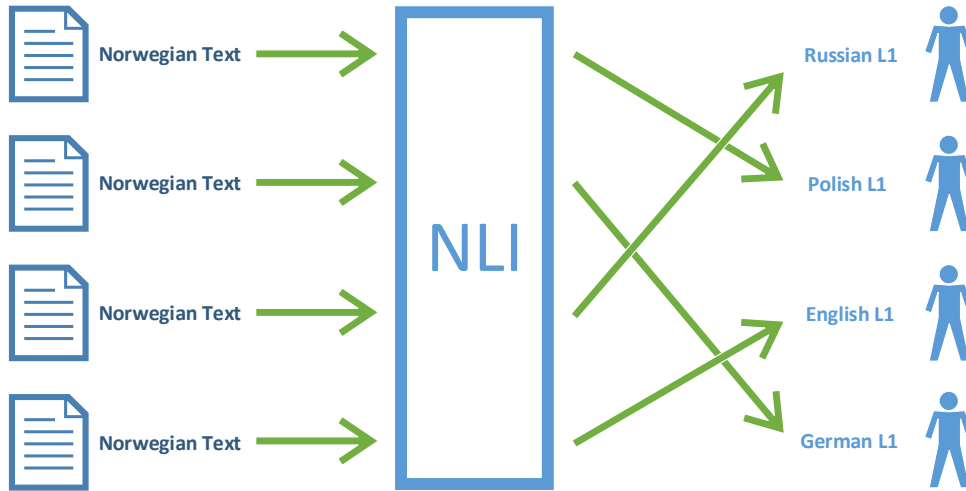[5]http://web.science.mq.edu.au/%7Esmalmasi/data/norwegian-funcwords.txt

Figure 2: Illustration of our NLI system that identifies the L1 of Norwegian learners from their writing.

lower accuracy, possibly due to data sparsity, so we do not include them.

We observe 328 different tags in the data resulting in 9k unique bigrams and 61k trigram features.

Mixed POS-Function Word n-grams Previously Wong et al. (2012) proposed the use of POS n-grams which retained the surface form of function words instead of using their POS tag. Example mixed trigrams include "`the NN that`" or "`NN that VBZ`". They demonstrated that such features can outperform their pure POS counterparts. Here we also use our above-described function word list to generate such mixed n-grams.

5.1 Results

The results for all of our features are shown in Table 2. We compare against a majority class baseline of 13% which is calculated by using the largest class, in this case Polish, as the default classification label chosen for all texts.

The distribution of function word unigrams and bigrams is highly discriminative, yielding accuracies of 51.1% and 50.0%, respectively. These are well-above the baseline and suggest the presence of L1-specific grammatical and lexical choice patterns that can help distinguish the L1, potentially due to cross-linguistic transfer. Such lexical transfer effects have been previously noted by researchers and linguists (Odlin, 1989). These effects are mediated not only by cognates and similarities in word forms, but also word semantics.

Feature	Accuracy (%)
Majority Baseline	13.0
Function Words	51.1
Function Word bigrams	50.0
Part-of-Speech unigrams	61.2
Part-of-Speech bigrams	66.5
Part-of-Speech trigrams	62.7
POS/Function Word trigrams	78.1
All features combined	**78.6**

Table 2: Norwegian Native Language Identification accuracy for the features used in this study.

The purely syntactic POS n-gram models are also very useful for this task, with the best accuracy of 66.5% for POS bigrams. This is the highest NLI accuracy achieved using POS n-grams. Using the 11-class TOEFL11 data, none of the shared task entries or subsequent studies have achieved accuracies of 60% or higher, with results usually falling in the 40–55% range. We also note that our POS n-gram performance plateaus with bigrams. This deviates from previous NLI results where trigrams usually yield the highest accuracy. This, alongside the higher accuracy, could potentially be a result of the tags being manually corrected by annotators, leading to more accurate tags and thus classification accuracy. However, this does not entirely explain why performance degrades when using trigrams. This could be due to tagset size and the number of features because with 328 tags, this is the largest tagset used for NLI to date.
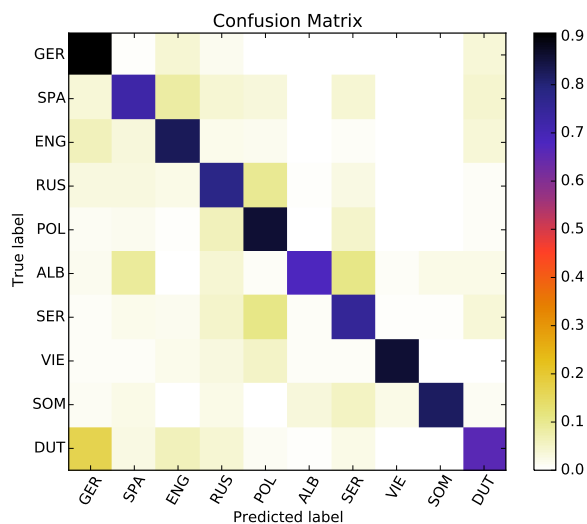
Figure 3: Confusion matrix for our 10 classes.

Figure 4: A learning curve for our Norwegian NLI system trained on all features.

The mixture of POS and function word n-grams provides the best result for a single feature with 78.1% accuracy. This is consistent with previous findings about this feature type.

Finally, combining all of the models into a single feature vector provides the highest accuracy of 78.6%, which is only slightly better than the best single feature type.

Figure 3 shows the normalized confusion matrix for our results. German and Polish are the most correctly classified L1s, while the highest confusion is between Dutch–German followed by Serbian–Polish and Russian-Polish. This is not surprising given that these pairs are from the same families: Germanic and Slavic. We were however surprised by the substantial confusion between Albanian and Spanish, even though the languages are not typologically related.

We also analyze the rate of learning for our classifier. A learning curve for a classifier trained on all features is shown in Figure 4. We observe that while there is a rapid initial increase in accuracy, performance begins to level off after around 1,500 training documents.

6 Identifying Non-Native Writing

Our second experiment involves using the above-described features to classify Norwegian texts as either Native or Non-Native. To achieve this we use 250 control texts we generated from the ASK Corpus that were written by native Norwegian speakers; these texts represent the Native class. This is contrasted against the Non-Native class
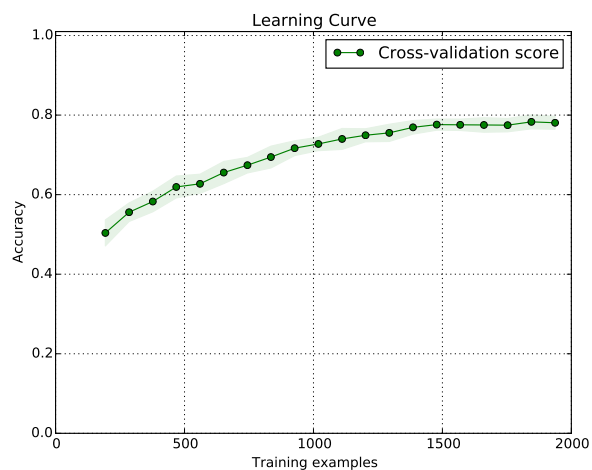
which includes 250 texts sampled from each language[6] listed in Table 1.

Feature	Accuracy (%)
Random Baseline	50.0
Function Words	90.0
Function Word bigrams	94.2
Part-of-Speech unigrams	95.0
Part-of-Speech bigrams	98.4
Part-of-Speech trigrams	98.5
POS/Function Word trigrams	98.6
All features combined	**98.8**

Table 3: Accuracy for classifying Norwegian texts as either Native or Non-Native.

6.1 Results

The results of our final experiment for distinguishing non-native writing are listed in Table 3. They demonstrate that these feature types are highly useful for discriminating between Native and non-Native writings, achieving 98.8% accuracy by using all feature types. POS/Function Word mixture trigrams are the best single feature in this experiment.

These results show that the language productions of native speakers are very different to those of learners, enabling our models to distinguish them with almost perfect accuracy.

---

[6]We sample evenly with 25 texts per non-native L1 class.

7 Discussion and Conclusion

We presented the first Norwegian NLI experiments, achieving high levels of accuracy that are comparable with previous results for English and other languages. A key objective here was to investigate the efficacy of syntactic features for Norwegian, a language which is different to English in some aspects such as morphological complexity. The features employed here could also identify non-native documents with 99% accuracy.

Another contribution of this work is the identification of a new dataset for NLI. Tasks focused on detecting L1-based language transfer effects – such as NLI – require copious amounts of data. Contrary to this requirement, researchers have long noted the paucity of suitable corpora[7] for this task (Brooke and Hirst, 2011). This is one of the research issues addressed by this work. The introduction of this corpus can assist researchers test and verify their methodology on multiple datasets and languages.

This study is also novel in its use of post-corrected POS tags. As noted in §5.1, while the POS-based results here are different from those of previous studies that have used automated tagging methods, it is unclear if this is due to the use of post-edited tags or the large size of the tagset. This is an issue that merits further investigation. Additional results from a fully experimental setup using multiple sets of automatic and gold standard POS tags for the same texts can help provide better insight here. The ASK corpus does not include the original POS tags obtained automatically using the Oslo-Bergen tagger prior to human editing and the texts would need to be re-annotated for such a study. This is left for future work.

There are a number of directions for future research. There have been a number of interesting NLI that could also be tested on this data. These include oracles for determining the upper-bound on classification accuracy (Malmasi et al., 2015), analyses of feature diversity and interaction (Malmasi and Cahill, 2015), and large-scale cross-corpus experiments (Malmasi and Dras, 2015b).

The application of more linguistically sophisticated features also warrants further investigation, but this is limited by the availability of Norwegian NLP tools and resources. For example, the use of a Norwegian constituency parser could be used to study the overall structure of grammatical constructions as captured by context-free grammar production rules (Wong and Dras, 2011). Another possible improvement is the use of classifier ensembles to improve classification accuracy. This has previously been applied to other classification tasks (Malmasi and Dras, 2015a) and English NLI (Tetreault et al., 2012) with good results.

References

Daniel Blanchard, Joel Tetreault, Derrick Higgins, Aoife Cahill, and Martin Chodorow. 2013. TOEFL11: A Corpus of Non-Native English. Technical report, Educational Testing Service.

Julian Brooke and Graeme Hirst. 2011. Native language detection with 'cheap' learner corpora. Presented at the *Conference of Learner Corpus Research*, University of Louvain, Belgium.

Julian Brooke and Graeme Hirst. 2012. Measuring interlanguage: Native language identification with L1-influence metrics. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 779–784, Istanbul, Turkey, May.

Ana Dıaz-Negrillo, Detmar Meurers, Salvador Valera, and Holger Wunsch. 2010. Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. In *Language Forum*, volume 36, pages 139–154.

Jan Terje Faarlund, Svein Lie, and Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Columbia University Press.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Sylviane Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. 2009. *International Corpus of Learner English (Version 2)*. Presses Universitaires de Louvain, Louvian-la-Neuve.

David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks. 2006. A Close Look at Skip-gram Modelling. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1222–1225, Genoa, Italy.

---

[7]An ideal NLI corpus should have multiple L1s, be balanced by topic, proficiency, texts per L1 and be large in size.

Einar Haugen. 2009. Danish, norwegian and swedish. In Bernard Comrie, editor, *The world's Major Languages*, pages 197–216. Routledge.

Scott Jarvis and Scott Crossley, editors. 2012. *Approaching Language Transfer Through Text Classification: Explorations in the Detection-based Approach*, volume 64. Multilingual Matters Limited, Bristol, UK.

Ekaterina Kochmar. 2011. Identification of a writer's native language by error analysis. Master's thesis, University of Cambridge.

Moshe Koppel, Jonathan Schler, and Shlomo Argamon. 2009. Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, 60(1):9–26.

Shervin Malmasi and Aoife Cahill. 2015. Measuring Feature Diversity in Native Language Identification. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 49–55, Denver, Colorado, June. Association for Computational Linguistics.

Shervin Malmasi and Mark Dras. 2014a. Arabic Native Language Identification. In *Proceedings of the Arabic Natural Language Processing Workshop (EMNLP 2014)*, pages 180–186, Doha, Qatar, October. Association for Computational Linguistics.

Shervin Malmasi and Mark Dras. 2014b. Chinese Native Language Identification. pages 95–99, Gothenburg, Sweden, April. Association for Computational Linguistics.

Shervin Malmasi and Mark Dras. 2014c. Finnish Native Language Identification. In *Proceedings of the Australasian Language Technology Workshop (ALTA)*, pages 139–144, Melbourne, Australia.

Shervin Malmasi and Mark Dras. 2014d. Language Transfer Hypotheses with Linear SVM Weights. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1385–1390, Doha, Qatar, October. Association for Computational Linguistics.

Shervin Malmasi and Mark Dras. 2015a. Language Identification using Classifier Ensembles. In *Proceedings of LT4VarDial - Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects*, Hissar, Bulgaria, September.

Shervin Malmasi and Mark Dras. 2015b. Large-scale Native Language Identification with Cross-Corpus Evaluation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2015)*, pages 1403–1409, Denver, CO, USA, June. Association for Computational Linguistics.

Shervin Malmasi, Sze-Meng Jojo Wong, and Mark Dras. 2013. NLI Shared Task 2013: MQ Submission. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 124–133, Atlanta, Georgia, June. Association for Computational Linguistics.

Shervin Malmasi, Joel Tetreault, and Mark Dras. 2015. Oracle and Human Baselines for Native Language Identification. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, Denver, Colorado, June. Association for Computational Linguistics.

Terence Odlin. 1989. *Language Transfer: Cross-linguistic Influence in Language Learning*. Cambridge University Press, Cambridge, UK.

Lourdes Ortega. 2009. *Understanding Second Language Acquisition*. Hodder Education, Oxford, UK.

Ria Perkins. 2014. *Linguistic identifiers of L1 Persian speakers writing in English: NLID for authorship analysis*. Ph.D. thesis, Aston University.

Ben Swanson and Eugene Charniak. 2014. Data driven language transfer hypotheses. *EACL 2014*, page 169.

Kari Tenfjord, Hilde Johansen, and Jon Erik Hagen. 2006a. The "Hows" and the "Whys" of Coding Categories in a Learner Corpus (or "How and Why an Error-Tagged Learner Corpus is not ipso facto One Big Comparative Fallacy"). *Rivista di psicolinguistica applicata*, 6(3):1000–1016.

Kari Tenfjord, Paul Meurer, and Knut Hofland. 2006b. The ASK corpus: A language learner corpus of Norwegian as a second language. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 1821–1824.

Kari Tenfjord, Paul Meurer, and Silje Ragnhildstveit. 2013. Norsk andrespråkskorpus - A corpus of Norwegian as a second language. In *Learner Corpus Research Conference (LCR 2013)*.

Joel Tetreault, Daniel Blanchard, Aoife Cahill, and Martin Chodorow. 2012. Native Tongues, Lost and Found: Resources and Empirical Evaluations in Native Language Identification. In *Proceedings of COLING 2012*, pages 2585–2602, Mumbai, India, December. The COLING 2012 Organizing Committee.

Joel Tetreault, Daniel Blanchard, and Aoife Cahill. 2013. A Report on the First Native Language Identification Shared Task. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 48–57, Atlanta, Georgia, June. Association for Computational Linguistics.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *IN PROCEEDINGS OF HLT-NAACL*, pages 252–259.

Bertus Van Rooy and Lande Schäfer. 2002. The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies*, 20(4):325–335.

Maolin Wang, Shervin Malmasi, and Mingxuan Huang. 2015. The Jinan Chinese Learner Corpus. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 118–123, Denver, Colorado, June. Association for Computational Linguistics.

Sze-Meng Jojo Wong and Mark Dras. 2011. Exploiting Parse Structures for Native Language Identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1600–1610, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Sze-Meng Jojo Wong, Mark Dras, and Mark Johnson. 2012. Exploring Adaptor Grammars for Native Language Identification. In *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pages 699–709.