WOVEN (Word-Order Variation Encoder): Improving Second Language Structure Comprehension through Machine Translation Explainability

Daniel Gallagher

Student ID: 18401492

Supervisor: Julie Berndsen

UCD School of Computer Science
University College Dublin

December 23, 2021

Chapter 1: **Project Description**

Language structure can be a major obstacle for learners in acquiring a second language (L2). This project strives to remedy some of these difficulties by detecting and expressing the variable word order in English-German translations in an intuitive and helpful way. In the final build, the application should take a sentence in the user's input language (English or German), translate it into the target language, analyse the translation for the variation in input-output word order formation, and lastly express this to the user in a way which will help them understand the differing structure.

A Neural Machine Translation (NMT) model is the first step in the development cycle. Marian (Junczys-Dowmunt et al. 2018) is an efficient and open-source NMT framework written in pure C++ and with minimal dependencies. Hugging Face provides a pre-trained Marian MT model which can be used and deployed in python. At this layer, the focus is solely on translation accuracy.

The second development phase will relate to the explanation layer. SHAP (Lundberg and Lee 2017) is a model-agnostic explainability framework that will be responsible for drawing the relations between input tokens and output tokens.

Here the focus shifts towards *explainability* or *interpretability*. That is, why did the model choose to make the structure of the translation so, and answering the question of which input elements were "responsible" for individual output elements (i.e which word in English corresponds to which in German). Note that this is inherently different from searching an English-German dictionary, as the explainer model is "aware" of which English elements are related to which German elements without any prerequisite knowledge of English or German and can account for the issues that arise in attempting to directly translate between the two. The word order variations between each language will be found and stored by WOVEN (Word-Order Variation Encoder), which will then be used in the next phase to communicate these variations to the learner.

The final phase develops a web-based application running on Flask. Flask is a web application micro-framework which sacrifices built-in functionality for a simpler code-base and reduced overhead. It is suitable for the scope of this project, adhering to the KISS principle of project design. This end of this phase marks the final of three in the project's development.

1.1 Background

The ease and speed of second-language acquisition can be greatly improved by the tools available to the learner. Translation services are widely available for language learners, and traditionally focus on taking an input A and outputting the correct translation B. However, receiving the most accurate and correct translation is only one piece of the puzzle to effective language learning. Oftentimes, the structure and formation of a second language is one of the most challenging hurdles to overcome in effective language learning.

Multilingual analysis has shown persistent variability (globally), flexibility (locally) and malleability in natural language word-order formation, with no exception for English and German (Östling 2015). Global variability can be demonstrated through the unique order of the subject, verb, and object in many languages when cross-compared. One often sees speakers move between both Subject-Verb-Object (SVO) and Verb-Subject-Object (VSO) in German, while adhering to SVO in English. One sees local flexibility through the allowance in language of sentences with the same meaning to be expressed in varying formations. In German, examples here would be “Der Junge gab dem Mädchen den Fußball”, “Dem Mädchen gab der Junge den Fußball”, and “Den Fußball gab der Junge dem Mädchen.” Each conveying the same meaning (though with varying emphasis), while flexing much looser word order rules than the English counterpart “The boy gave the football to the girl.” Lastly, malleability expresses the ability of word order to change over time. This can be shown through the German speaker's gradually shifting verb placement in some relative clauses following conjunctions such as “weil” (Eng. “because”). (Farrar 1999)

Therefore, there is an evident need here for tools that provide the learner with a better understanding of the *why* behind the formation of a translation, rather than a focus purely on translation accuracy.

1.2 Related Work

Deep neural networks focus on optimising a vast amount of parameters in an attempt to model more and more complex datasets. Each node within a neural network will optimise an activation function (whether that be ReLU, SoftPlus, Sigmoid, etc.) using weights and biases. Backpropogation is used to relay new information about the structure of the system back to these parameters. Expanding on these core concepts, we can model increasingly complex and abstract systems. Long Short-Term Memory networks (LSTM) (Sundermeyer, Schlüter, and Ney 2012) are a form of Recurrent Neural Network (RNN) that show an impressive ability to handle long-term dependencies in texts that are heavily context-dependent. They have a “loop”-like structure which relays information about the text as a whole, so as to “remember” important pieces of context.

A major development in NMT has been a research shift towards the concept of “attention”. Attention (Vaswani et al. 2017) can be described as placing more emphasis (or weight) on elements of input that contribute the most to the entirety or particular elements of the output. This is the result of a particularly effective model released in 2017 called a transformer. This is an encoder-decoder architecture which dispenses with recurrence and convolutions entirely, incorporating attention as the sole mechanism. The transformer model achieved a 28.4 BLEU score on the WMT 2014 English-to-German translation task, showing an improvement of over 2 BLEU on the prior best. There are many such transformers in widespread use for MT tasks currently, with one such open-source framework being the aforementioned Marian.

A tension between accuracy and interpretability has been made evident in recent years to the machine learning community. (Lundberg and Lee 2017) As models continue to expand in size and complexity, the ability to associate any explanations or reasoning behind predictions has become an increasingly difficult and important task. There are, however, many such methods of extracting explainability from models, with research being driven by the increasing need for model interpretability. SHAP is one such explainability framework, focusing on interpreting predictions by assigning each feature an “importance value” for a particular prediction. That is, how much did any given input feature contribute to the final prediction. This approach introduces a view of any model's prediction as a model itself, generally referred to as an explanation model. SHAP has demonstrated a strong ability to interpret the results of NMT models, however to the best of the author's knowledge, there has not yet been an NMT project developed with the sole intention of improving L2 learner's language structure comprehension.

1.3 Resources and Datasets

None required.

Chapter 2: Related Work & Ideas

The literature for approaches to language learning and their continuous improvement is rich. In section 1.2, an introductory overview of recent developments in NMT and XAI was given, however this review aims to give a more broad overview of language learning and the potential for these developments to aid the learner in the classroom.

Section 2.1 introduces the thought and research surrounding L2 learning since the early 20th century. The introduction of technological improvements has further shifted the landscape and this will be examined in section 2.2, with the move to completely online learning in a pandemic being explored in section 2.3. A general overview of student difficulties with regard to L2 learning is explored in section 2.4.

A more specific focus on machine translation, its progress, and its potential improvements in the context of L2 learning is traversed in section 2.5. Lastly, the more recent developments in extracting explanations from machine learning models are summarised in section 2.6, with a focus on their potential to improve L2 structure comprehension.

2.1 General Trends in Language Learning

Much of the most important research and thought surrounding the learning of foreign languages has been developed within the last one hundred years. (Otto 2017) This relatively recent focus on pedagogical methodology in the classroom has reinvented the approach of both student and teacher to language learning multiple times over, with major developments lying in the technology we use. While the essential media used in the process have remained constant - that is written texts, drawings, photos, audio, and video - the way in which we deliver them has changed dramatically.

The Grammar-Translation method was the traditional approach used commonly in the early 20th century. This method emphasised a grammar-first approach, stemming from the teaching of Latin and Greek. Other methods such as the Natural and Direct method stressed the importance of oral practice in *second language acquisition* (SLA). Beginning mid-century, repetition in SLA began to be heavily promoted as an important part of L2 recall, shifting the focus of language teaching.

Throughout the first half of the twentieth century, technology in language learning as we know it was essentially non-existent bar the possibility of playing audio. It wouldn't be until the late 1960s and early 1970s that technology such as the computer could begin to show potential in the language learning domain, termed *Computer-Assisted Language Learning* (CALL). CALL allowed assignments and worksheets to be automatically corrected, providing students with immediate feedback and teachers with the opportunity to focus on feedback rather than correction.

An important emphasis has been placed upon the distinction between *tool* and *tutorial* CALL, with *tool* CALL referring to applications that can be used independently by students for e-learning, and *tutorial* CALL referring to programs which are generally used by the teachers and students in the classroom.

CALL has moved from a niche area of education integrated with computer science to a primary component of the teaching classroom. Software, web applications, and language tools have been in continuous development and each year improve the classroom experience for language learners.

In recent times we have seen them become the classroom themselves, as language classes move online in the advent of a worldwide pandemic.

2.2 The Introduction of Online Language Learning

From the mid-1980s to the present day, language learning has been continuously evolving due to the introduction of online tools into the classroom and the wide availability of computing devices for students and teachers alike. The internet has brought with it a host of translation tools, online dictionaries, e-learning applications and user forums, all of which can be freely accessed by the learner from their home or classroom.

Online Language Learning (OLL) generally refers to a web-facilitated class, a blended or hybrid course, or a fully virtual online class (Blake 2011), and has seen a large increase in popularity from 2000 onwards. Blake shows that post-secondary online course enrollment as a percentage of total enrollment has steadily increased year-by-year from 2002 to 2008, growing from 9.6% to 25.3%. In the recent decade, the number of people signing up for at least one online course has averaged an increase of 19% per year.

The perspective of the language learner also changed during this time, and this has pushed for more independent, self-motivated learning in accompaniment to the traditional classroom. At the mark of the millennium, pedagogical research shifted to view the learner as a social being, in which SLA occurs through social interactions mediated by the target language. A further emphasis was placed on the *L2 motivational self system* (Dörnyei 2019), in which the learner is motivated by the *ideal* L2 self, *ought-to* L2 self, and the L2 *learning experience*. This final element, the learning experience, has been highlighted as the most powerful predictor of motivated behaviour.

These key insights have been a large part of a shift towards online, independent learning tools. Mobile applications such as *Duolingo* and *Babbel* have developed platforms on which the learner can improve their target language without the need of an experienced speaker present. Additionally, online dictionaries such as *focloir.ie* or *teanglann.ie* provide easy access to translation pairs for Irish or English, optimising the age-old process of searching for a word in a dictionary. Translation tools such as DeepL and Google Translate have also become extremely powerful for popular languages such as English, Mandarin, German, Spanish, and French. *Rosetta Stone*, a major CALL-based language learning company founded in 1992, has even developed an e-learning Duolingo-style mobile application to innovate on its classic language learning package.

As (Blake 2011) notes, even traditional publishers of second language text- and work-books have begun to provide accompanying online grammar exercises as supplement material. Though these have been shown to offer students a relatively low level of interactivity and a restricted ability to construct meaning independently.

In more recent years, the term *Distance Learning* (DL) has become increasingly popular, referring to the portion of OLL which remains fully online and "at-distance". In his book *Brave New Digital Classroom* (Blake 2008), Blake has found that overall, the improvement in the language abilities of students who participated in fully online courses was found to be greater than that of students in traditional face-to-face classes. However, it is not immediately clear why this may be the case, and there is still much research which remains to be completed here.

2.3 Results of Online Language Teaching During the COVID-19 Pandemic

Though it will be many years before the full effect of the move to online teaching during the COVID-19 pandemic will be known, it is immediately clear that the trend towards increased DL has been dramatically accelerated. Prior to the pandemic, much of online teaching took place on *Massive Open Online Courses* (MOOCs). These could generally be completed in a student's own time, and did not have the same level of peer interaction as with traditional courses. Only some academic institutions offered online classes that were run in the traditional manner yet through an online medium. This changed rapidly as restrictions were brought in on face-to-face teaching and learning in March of 2020.

In the context of Ireland, the pandemic moved all primary-, secondary- and third-level education online, with mass-deferral not an option for most institutions. Any preconception that language learning must be face-to-face was immediately put to the test, as teachers and students were forced to make a rapid transition to the online classroom.

Recent research conducted by (Maican and Cocoradă 2021) has shown that both positive and negative emotions are expressed by L2 students in their new virtual classroom environment during the pandemic. In a study conducted with over 20,000 students in Romania, 55.6% of students said E-learning should not replace face-to-face learning at all. However, just under half supported the use of e-learning platforms to some extent replacing traditional face-to-face learning.

With regard to language learning, students reported negative emotions generated by the lack of interaction with peers and teachers, yet positive emotions associated with retrospective enjoyment of the class, as measured by the *Foreign Language Enjoyment* (FLE) scale. It is likely that lower-achieving L2 learners were more likely to be at a disadvantage with online learning, as the face-to-face group environment provided a particularly needed engagement that online learning could not. It is likely that further research will reveal the extent to which certain groups of students are disadvantaged by online learning in the coming years.

2.4 Common Student Difficulties in Language Learning

2.4.1 Phonetic & Pronunciation Difficulties

One of the most common problems for students of English is its inconsistent phonetics. In German, the spelling system has been standardised so that words will almost always sound exactly as they are written, that is, they are phonetically consistent. In English, words such as "though" and "rough" can have the same ending but be pronounced completely differently.

Additionally, learners not only have problems with phonetic inconsistencies, but also with the actual difficulty of sound pronunciation in learning L2 coming from an L1. The more distant the "phonetic alphabet" of your L1 is from your L2, the more likely it is that you will find pronunciation particularly difficult. This is the reason L2 learners of Japanese find such difficulty coming from English (Ohata 2004). English learners of Spanish have particular difficulty with the notable "rolled r", a sound very different to any in English.

In the mid-20th-century, researchers put forward the *Contrastive Analysis Hypothesis* (CAH) (Munro 2018), in which the L2 classroom should take into account the L1 of learners in order to predict likely future pronunciation difficulties. The idea of having an "error inventory" for learners was understandable and quite appealing, though the exhaustive phonetic analysis that was sought after for learners never emerged as a useful classroom tool.

2.4.2 Structural & Semantic Difficulties

English and Korean come from different language families, and correspondingly take a considerable amount of effort on the part of the learner to learn one as an L2, when coming from the other as an L1. These differences are very evident in the syntactic structure and semantics of each.

In studying Korean students of English, (Cho 2004) notes that learners had particular problems with the structure and formation of the English language. For example, relative clauses will generally come prior to the nouns that they are modifying. In Korean, one would say *"I saw yesterday the man"* instead of *"the man I saw yesterday."* Korean has neither conjugation nor inflection, therefore verbs will not change as they do in English depending on the subject. Learners often have trouble completing the subject-verb agreement necessary in English.

L2 word order can present further challenges for students, especially when it is significantly altered from L1. It is generally accepted that German has *subject-object* (SO) word order, however often *object-subject* (OS) order also occurs (Bader and Häussler 2010). This can be for multiple reasons, for instance the use of an accusative or dative object as a verb argument. These subtle differences in a language can often leave the L2 language learner frustrated.

Students also have difficulty with *native speaker intuition*, that is, the natural understanding of when to use what even if not immediately obvious. For instance, students are often confused as to when to use the *definite*, *indefinite*, or no article before a noun. These problems persist on the part of the learner, and students often must immerse themselves completely in the language before they can advance to their desired level.

2.5 Translation Tools in the Classroom

2.5.1 Early Concerns

From the beginning of the millennium until 2015, translation tools saw a large growth in usage by L2 learners, while still remaining in their infancy of development for many languages. Languages such as Basque and Irish are two such examples, where a reliance on a relatively small amount of available data coupled with a smaller amount of investment from translation companies led to poor overall accuracy. The high likelihood of providing students with flawed and inexact translations was a persistent concern.

The attitude of teachers towards *Web-based Machine Translation* (WBMT) tools in the classroom has often been dismissive, often disregarding them as used by those who are lazy (Van Praag and Sanchez 2015). This is an understandable concern, connected to a worry that students will disengage from the language learning process and find the shortest link to their L1. It has indeed been shown that WBMT tools have been more likely to be used by struggling students than those who aren't (Briggs 2018). However, removing these tools entirely from the language learning process has been found to be ineffective in improving student's SLA ability, and potentially places struggling students at major disadvantage.

The combination of these two primary worries among teachers, (i) *misleading and inaccurate translations* and (ii) *student disengagement*, has often held enthusiasm back among pedagogical circles in the use of MT tools in L2 learning.

2.5.2 Addressing Concerns & More Recent Developments

A major recent development in MT has been that of Neural Machine Translation (NMT) (Wu et al. 2016), first gaining widespread recognition through Google in 2016. The proof was in the pudding, as English-French translations were put to the test with these new methods and found to reveal remarkable improvements in translation accuracy (Briggs 2018) when compared to the preferred methods of the time such as *Phrase-based Statistical Machine Translation*(PBSMT). This shifted the focus within the MT landscape and has been the springboard for major accuracy improvements in recent years.

This has directly tackled concern (i) from the end of section 2.5.1, that is, misleading and inaccurate translations. This is especially the case among more closely related and resource-abundant languages such as Spanish and French. MT has become much better than it was before, yet is still flawed. If there is much underlying meaning and connotation present, it becomes more difficult for MT to produce reliable results. However, for learners from level A1 to B1 on the Common European Framework of Reference for Languages (CEFR) scale, that is, beginner to lower intermediate, difficult translations of this kind are unlikely to be as common as simpler, more reliable ones.

With many students able to access more reliable translations without the presence of a teacher or native speaker, researchers have also attempted to answer whether translation tools are enhancing or harming learning. (Clifford, Merschel, and Munné 2013) found in a survey conducted among L2 students that only 12% *never* use MT in the learning process, with the majority believing that it was at least sometimes of benefit. When asked in what ways they felt it was beneficial, common answers were *increasing vocabulary* (85%), *increasing grammatical accuracy* (47%), and *building confidence* (32%). Interestingly, the more advanced learners found MT to be of more use than the intermediate learners, and the intermediate more-so than the beginners. This suggests that MT acts as a strong tool for learning reinforcement and a weaker tool for teaching. Overall, the majority of students studied find that MT aids in L2 learning, and preconceptions of MT use as "lazy" are misguided.

2.5.3 Gaps in the Research

In assessing the quality of translation tools, the focus has been predominantly on the translation accuracy measured by BLEU scores. It is clear that this metric is incredibly important, but from a language learner's perspective, it's not everything. Tools such as DeepL which provide the learner with a definition, noun article, and example prove to have significant advantages for the learner over those which don't, regardless of translation accuracy. There is more to learning than precision.

Perhaps more time should be invested in considering further metrics, those which can take into account the person often behind the screen: an L2 learner. How well did the translation tool help define each word? Did it offer much in the way of explanation for word-order variation between L1 and L2? Why is the dative and not accusative case used? Further research in the field of *Explainable AI* (XAI) may bring help alleviate these questions for students using translation tools in L2 learning.

2.6 Explainable AI (XAI)

XAI is the ability to express an interpretation of why a model has chosen a particular output or made a particular decision. In less complex machine learning methods such as regression, this is quite simple. We can see transparently through every step in the process, and know what is determining the regression line. For data of higher dimensions and models of higher complexity, this becomes far less trivial. It is possible that with further research, translation tools can begin to examine the potential advantages of XAI in their models. The potential learning benefits of the L2 learner being given "interpretations" for the structure, semantics, and morphology of a language are distinctly apparent. While the focus until now has been primarily on improving accuracy, the field must reflect on who it has built this technology for, and in what way can it be made more suitable for their purposes.

2.6.1 SHAP Explainability Framework

(Lundberg and Lee 2017) published their influential paper "A Unified Approach to Interpreting Model Predictions" in 2017, which has since become influential in the field of model explainability. Their approach takes the application of Shapley values from game theory (Shapley 1953) and applies it to model explainability. This method, entitled SHAP, draws a unique comparison between players of a game and features in model input, resulting in an effective method for measuring their effect on model output. Features are substituted in and out of the input in order to measure the effect on the model's output, as players would be brought in and out of the game. Though expensive, this has proven an effective method of investigating which subsets of the input are most responsible for subsets of the output.

2.6.2 Derivatives

There has been the publication of many derivatives and extensions of SHAP in the literature since (Lundberg and Lee 2017). Many adjust the algorithm used to calculate the Shapley values and hence the feature contributions, while others alter the assumptions made before calculating these contributions.

The *Asymmetric Shapley Values* (ASV) method (Frye, Rowat, and Feige 2019) relaxes the assumption of feature symmetry in the input, having the effect of taking into account causal relationships between input features. That is, if feature A has increases the likelihood of feature B occurring, feature A will take more 'credit' for the final prediction output. *NeuronShapley* (Ghorbani and Zou 2021) on the other hand concerns itself with isolating specific subsets of nodes or units in a deep-learning model which have the greatest contribution to the output. *CausalShapley* (Heskes et al. 2020), *SealSHAP* (Parvez and Chang 2021), and *ShapleyFlow* (Wang, Wiens, and Lundberg 2021) are three additional recent variations of SHAP-based explainability.

2.6.3 Application in Machine Translation

The application of SHAP and its derivatives to *Natural Language Processing* (NLP) has not yet been extensively studied. NLP can present difficulties in the string-type input that must be fed to deep-learning models. Word embeddings and variable input length are two such problems. However, documentation of the successful implementation of SHAP to MT models does exist (Lundberg 2018a).

In an MT context, this allows us to investigate which individual words, or group of words, are most responsible for those in one's target language. English-Spanish and English-French translations have been successfully demonstrated, however there remains many gaps in the research for the further application of SHAP to MT beyond this.

WOVEN directly explores this research space in exploring the practical application of SHAP to MT, using the resulting feature contributions to provide insight to the learner of the varying word order present in English-German translations. Figure 2.1 provides a demonstration of how this may be presented to the learner once in use. Three examples of ways in which word-order variation occurs are presented: *emphasis*, *inversion*, and the presence of *modal verbs*. With the development of this project, the space for research in the practical application of SHAP to MT can be finally explored, and the potential lying behind it revealed.

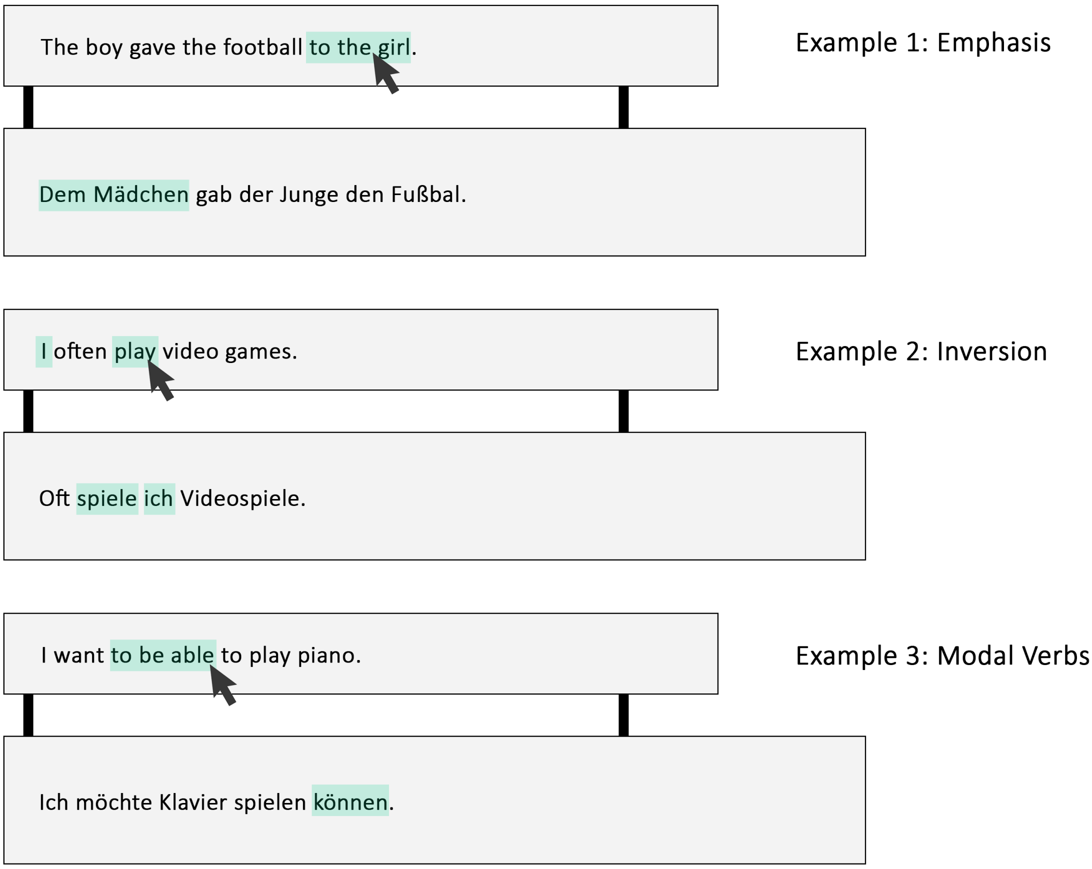

Figure 2.1: Potential Word-Order Variations revealed through SHAP-based Explainability

Chapter 3: **Project Management Plan**

Thorough project planning is an essential component to the development of any application. This project spans four development phases: (i) *Off-the-shelf Model Evaluation*, (ii) *Translation Word-Order Variation*, (iii) *Front-End User Interface*, and (iv) *Completion*.

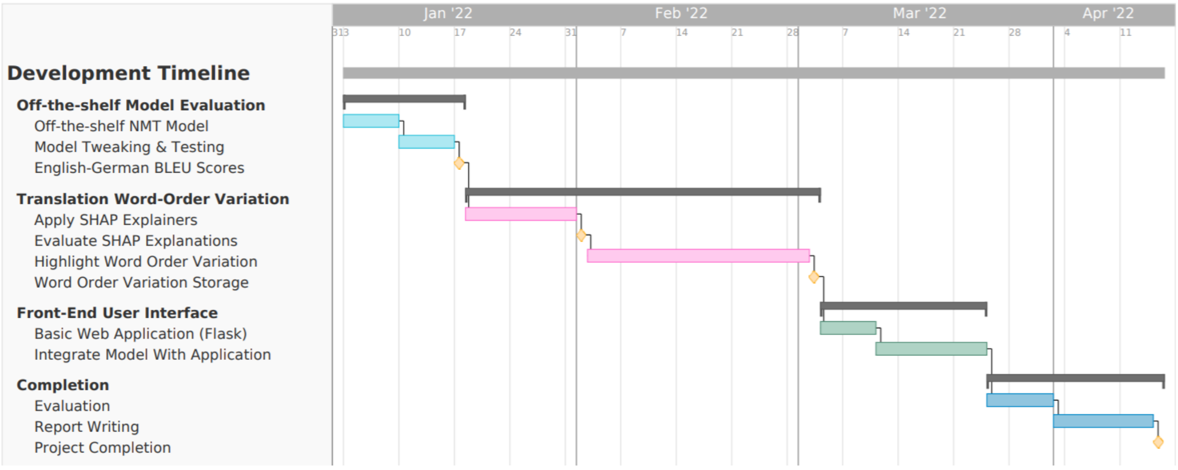

Figure 3.1: Project Development Gantt Chart

The first phase consists of using an off-the-shelf pre-trained NMT model for English-German translations using (Junczys-Dowmunt et al. 2018). Tweaking and testing is necessary in order to ensure that the model meets a high standard of translation, however no training will be required which will simplify this phase significantly.

The second phase consists of using the SHAP python library (Lundberg 2018b). Having worked extensively with SHAP during my semester at the Technical University of Munich, I believe it contains a huge amount of potential in revealing the connections between word-order structures between two languages, hence making translations further comprehensible to L2 learners. This phase will be the most important, and is therefore afforded a significant period of time during February where it can be focused on. The primary goal is to build a tool entitled WOVEN, which finds the structural links between L1 and L2 using SHAP interpretations and encodes these connections. These can be used by the learner to improve L2 structure comprehension. The ability of the tool to find these word-order variations given a set of test translations from our NMT model will mark the success of this phase and a significant milestone.

The third phase focuses on integrating WOVEN into a web application, in which users can translate sentences with the purpose of understanding how they change in structure from one language to the next (in this case, English-German). A basic user interface is all that's required for the scope of this project. I have chosen Flask over Django as the development framework in order to prioritise rapid development over extensive functionality. Approaching mid-March, this web application should be ready for use locally.

The final phase includes the final elements of the project. An ideal evaluation plan will be developed, where details are laid out of how one could appropriately test the tool with L2 learner feedback on how it contributed to their comprehension of language structure. This evaluation plan coupled with the project report is given a suitable three weeks for completion.

Bibliography

Bader, Markus and Jana Häussler (2010). "Word order in German: A corpus study". In: *Lingua* 120.3. Exploring the Left Periphery, pp. 717–762. ISSN: 0024-3841. DOI: https://doi.org/10.1016/j.lingua.2009.05.007. URL: https://www.sciencedirect.com/science/article/pii/S0024384109001466.

Blake, Robert (Jan. 2008). "Brave New Digital Classroom: Technology and Foreign Language Learning". In:

– (2011). "Current Trends in Online Language Learning". In: *Annual Review of Applied Linguistics* 31, pp. 19–35. DOI: 10.1017/S026719051100002X.

Briggs, Neil (Apr. 2018). "Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses". In: *JALT CALL Journal* 14, pp. 3–24. DOI: 10.29140/jaltcall.v14n1.221.

Cho, Byung-eun (2004). *Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. The East Asian Learner*.

Clifford, Joan, Lisa Merschel, and Joan Munné (2013). "Surveying the Landscape: What is the Role of Machine Translation in Language Learning?" In:

Dörnyei, Zoltán (Mar. 2019). "Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System". In: *Studies in Second Language Learning and Teaching* 9, pp. 19–30. DOI: 10.14746/ssllt.2019.9.1.2.

Farrar, Kimberley (1999). "Explanations for Word Order Change in Modern German". In: *Zeitschrift für Dialektologie und Linguistik* 66.1, pp. 1–30. ISSN: 00441449. URL: http://www.jstor.org/stable/40504445.

Frye, Christopher, Colin Rowat, and Ilya Feige (2019). "Asymmetric Shapley values: incorporating causal knowledge into model-agnostic explainability". In: *NeurIPS 2020*. URL: https://arxiv.org/pdf/1910.06358.pdf.

Ghorbani, Amirata and James Zou (2021). "Neuron Shapley: Discovering the Responsible Neurons". In: *NeurIPS 2021*. URL: https://proceedings.neurips.cc//paper/2020/file/41c542dfe6e4fc3deb251d64cf6ed2e4-Paper.pdf.

Heskes, Tom et al. (2020). "Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models". In: *NeurIPS 2020*. URL: https://proceedings.neurips.cc//paper/2020/file/32e54441e6382a7fbacbbbaf3c450059-Paper.pdf.

Junczys-Dowmunt, Marcin et al. (July 2018). "Marian: Fast Neural Machine Translation in C++". In: *Proceedings of ACL 2018, System Demonstrations*. Melbourne, Australia: Association for Computational Linguistics, pp. 116–121. DOI: 10.18653/v1/P18-4020. URL: https://aclanthology.org/P18-4020.

Lundberg, Scott (2018a). "Machine Translation Explanations". In: URL: https://shap.readthedocs.io/en/latest/example_notebooks/text_examples/translation/Machine%5C%20Translation%5C%20Explanations.html#English-to-French.

– (2018b). "Welcome to SHAP Documentation". In: URL: https://shap.readthedocs.io/en/latest/.

Lundberg, Scott and Su-In Lee (2017). *A Unified Approach to Interpreting Model Predictions*. arXiv: 1705.07874 [cs.AI].

Maican, Maria-Anca and Elena Cocoradă (2021). "Online Foreign Language Learning in Higher Education and Its Correlates during the COVID-19 Pandemic". In: *Sustainability* 13.2. ISSN: 2071-1050. DOI: 10.3390/su13020781. URL: https://www.mdpi.com/2071-1050/13/2/781.

Munro, Murray J. (2018). "How Well Can We Predict Second Language Learners' Pronunciation Difficulties?." In: *The CATESOL Journal* 30, pp. 267–281.

Ohata, Kota (Jan. 2004). "Title: Phonological Differences between Japanese and English: Several Potentially Problematic Areas of Pronunciation for Japanese ESL/EFL Learners". In: *Asian EFL Journal* 6.
Östling, Robert (July 2015). "Word Order Typology through Multilingual Word Alignment". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Beijing, China: Association for Computational Linguistics, pp. 205–211. DOI: 10.3115/v1/P15-2034. URL: https://aclanthology.org/P15-2034.
Otto, Sue E. K. (2017). "From Past to Present". In: *The Handbook of Technology and Second Language Teaching and Learning*. John Wiley Sons, Ltd. Chap. 2, pp. 10–25. ISBN: 9781118914069. DOI: https://doi.org/10.1002/9781118914069.ch2. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118914069.ch2. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118914069.ch2.
Parvez, Md Rizwan and Kai-Wei Chang (2021). "Evaluating the Values of Sources in Transfer Learning". In: *NAACL 2021*. URL: https://aclanthology.org/2021.naacl-main.402.pdf.
Shapley, Lloyd S (1953). "A value for n-person games". In: *Contributions to the Theory of Games 2.28*, pp. 307–317.
Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney (2012). "LSTM Neural Networks for Language Modeling". In: *INTERSPEECH*.
Van Praag, Benjamin and Hugo Santiago Sanchez (2015). "Mobile technology in second language classrooms: Insights into its uses, pedagogical implications, and teacher beliefs". In: *ReCALL* 27.3, pp. 288–303. DOI: 10.1017/S0958344015000075.
Vaswani, Ashish et al. (2017). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].
Wang, Jiaxuan, Jenna Wiens, and Scott Lundberg (2021). "Shapley Flow: A Graph-based Approach to Interpreting Model Predictions". In: *AISTATS 2021*. URL: http://proceedings.mlr.press/v130/wang21b/wang21b.pdf.
Wu, Yonghui et al. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. arXiv: 1609.08144 [cs.CL].