NOMINALIZATION AND ARGUMENT STRUCTURE: AN EXPERIMENT WITH THE NOMLEX DATABASE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

MELIS KIZILDEMIR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

AUGUST 2022

Nominalization and Argument Structure: An Experiment with the NOMLEX Database

submitted by **MELIS KIZILDEMIR** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek
Dean, **Graduate School of Informatics** ________________

Dr. Ceyhan Temürcü
Head of Department, **Cognitive Science** ________________

Prof. Dr. Cem Bozşahin
Supervisor, **Cognitive Science, Middle East Technical University** ________________

Examining Committee Members:

Prof. Dr. Cem Bozşahin
Cognitive Science, Middle East Technical University ________________

Assoc. Prof. Dr. Barbaros Yet
Cognitive Science, Middle East Technical University ________________

Assist. Prof. Dr. Özkan Aslan
Faculty of Engineering, Afyon Kocatepe University ________________

Date: 31.08.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: **Melis Kızıldemir**

Signature :

ABSTRACT

NOMINALIZATION AND ARGUMENT STRUCTURE: AN EXPERIMENT WITH THE NOMLEX DATABASE

Kızıldemir, Melis
M.S., Department of Cognitive Science
Supervisor: Prof. Dr. Cem Bozşahin

August 2022, 63 pages

This study presents a method for predicting syntactic structures of deverbal event nominals that are derived from verbs used in their transitive meaning. The study takes a data-driven approach and fine-tunes pre-trained deep learning models. The problem is treated as a 15 class multi-label classification task and NOMLEX is used to create the classes. In order to leverage the language models learned by pre-trained models, sentences available in corpora which include the verb that the deverbal nominal is derived from is used to fine-tune the models. As a result of this, selectional preferences of verbs are provided to the model implicitly. Sentences where the verb is used in its transitive meaning are filtered from the corpora using EasyCCG CCG parser. The DeBERTa model scores 66.4% in sample based accuracy and 77.9% in sample based F_1 score, and the RoBERTa model scores 75.0% in label based F_1 score. When the models are evaluated on classes, at least one model performs better than baseline in 10 out of 15 classes. Deverbal nominal syntactic structures that only realize an argument in the possessive determiner position, two out of three syntactic structures that realize the subject in the posessive determiner position and two out of three syntactic structures that realize the subject in the noun modifier position are among unsuccessfully learned classes. It is concluded that determiners of an argument's realization in a syntactic position is dependent on both the argument and the syntactic position in question along with realization of other arguments. The presented method was able to learn these determiners to various extents.

Keywords: nominalization, deep learning

ÖZ

ADLAŞTIRMA VE ARGÜMAN YAPISI: NOMLEX VERİTABANI İLE BİR ÇALIŞMA

Kızıldemir, Melis
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Prof. Dr. Cem Bozşahin

Ağustos 2022, 63 sayfa

Bu çalışmada geçişli anlamlarında kullanılan fiillerden üretilen olay adlaştırmalarının sözdizimsel yapılarını tahmin etmek için bir yöntem sunulmaktadır. Çalışmada veri odaklı bir yaklaşım kullanmakta ve önceden eğitilmiş derin öğrenme modellerine ince ayar yapılmaktadır. Problem 15 sınıflı bir çok etiketli sınıflandırma işi olarak tanımlanmış ve sınıfların yaratılmasında NOMLEX kullanılmıştır. Önceden eğitilmiş derin öğrenme modellerinin bilgilerinden yararlanabilmek için modellerin ince ayarlamalarında korporada bulunan ve olay adlaştırmalarının üretildiği fiileri barındıran cümleler kullanılmıştır. Bunun bir sonucu olarak fiillerin seçici tercihleri modellere dolaylı bir şekilde sağlanmıştır. Fiillerin geçişli anlamlarında kullanıldığı cümlelerin korporadan filtrelenmesinde EasyCCG CCG ayrıştırıcı kullanılmıştır. DeBERTa tabanlı model örnek temelli doğrulukta %66.4 ve örnek temelli F_1 puanında %77.9 performans göstermiş, RoBERTa tabanlı model sınıf temelli F_1 puanında %75.0 performans göstermiştir. Modeller sınıflar üzerinde değerlendirildiğinde 15 sınıfın 10'unda en az bir model F_1 puanında referanstan daha iyi performans göstermiştir. Sadece iyelik belirteci bulunan olay adlaştırmalarının sözdizimleri, öznenin iyelik belirteci posizyonunda bulunduğu olay adlaştırmalarının sözdizimlerinin 3'te 2'si ve öznenin niteleyici pozisyonda bulunduğu olay adlaştırmalarının sözdizimlerinin 3'te 2'si başarısız olarak öğrenilen sınıflar arasındadır. Bir argümanın sözdizimsel bir pozisyonda bulunmasına sebep olan faktörlerin argümana, söz konusu pozisyona ve diğer pozisyonlarda bulunan diğer argümanlara bağlı olduğu sonucuna varılmıştır. Sunulan yöntem bu faktörleri çeşitli derecelerde öğrenmiştir.

Anahtar Kelimeler: adlaştırma, derin öğrenme

To my family

ACKNOWLEDGMENTS

I would like to express my gratitude to my family for their unconditional love and support.

I would like to thank my advisor Prof. Dr. Cem Bozşahin for the seemingly never ending knowledge he provided me throughout the study. This study wouldn't have been possible without his guidance and encouragement.

I would like to thank Kayra for his insights and loving company.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
CCG	Combinatory Categorical Grammar
ROC	Receiver Operator Characteristics
AUC	Area Under Curve
PR	Precision vs. Recall
BERT	Bidirectional Encoder Representations from Transformers
RoBERTa	Robustly Optimized BERT Approach
DeBERTa	Decoding-enhanced BERT with Disentangled Attention

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nominalization is the process of type-shifting a lexical item which is not a noun, to become the head of a noun phrase. Products of this process are called derived nominals.

A subset of derived nominals are gerundive nominals. They are quite predictable morphologically, semantically and syntactically (Chomsky, 1970):

(1) a. John is eager to please. $\rightarrow$ John's being eager to please

b. John is easy to please. $\rightarrow$ John's being easy to please

c. John refused the offer. $\rightarrow$ John's refusing the offer

d. John critized the book. $\rightarrow$ John's criticizing the book

In contrast to gerundive nominals are deverbal nominals. They take arguments like verbs and differ from gerundive nominals by having noun phrase internal structures. Unlike gerundive nominals they seem unpredictable in nature (Chomsky, 1970).

Deverbal nominals may or may not receive a derivational suffix. If they do, there are many choices for the nominalizer suffixes:

(2) a. I doubt that he will be on time. $\rightarrow$ my doubt that he will be on time

b. I donated blood. $\rightarrow$ the blood donation

c. He attended to the conference. $\rightarrow$ the attendance to the conference by him

They can semantically relate to their verbal counterparts in various ways:

(3) a. He taught physics to my niece. → My niece's physics teacher (the nominal incorporates the subject role)

b. He arrived early. → His early arrival (the nominal expresses the event of the verb)

Their syntax shows significant variability, even when verbal syntactic environment is controlled:

(4) a. He announced the product. → His announcement

b. Carthage destroyed Rome. → *Carthage's destruction, Carthage's destruction of Rome (subject cannot occur alone in possessive determiner position)

In the context of seemingly unpredictable morphology, semantics and syntax of deverbal nominals, this work focuses on understanding and predicting their syntactic structure. NOMLEX (Macleod et al., 1998) is a lexicon of deverbal nominals of English, and provides comprehensive information about the circumstances that can lead to nominalization of a verb and syntactic structures of subsequently formed deverbal nominals. This information enables a data-driven approach, which is taken in the study.

1.2 The Research Question

The syntactic structure of the sentence which the verb is present in is crucial in determining nominal syntactic structure. It is observed that verbs can produce different deverbal nominals according to the syntactic structures of sentences they are in:

(5) a. I adopted Snuffles. → Snuffles adoption.

b. I adopted Snuffles from shelter. → *Snuffles adoption from shelter.

However, when verbal syntactic environment is controlled there is still significant variability in nominal syntactic structure. In other words, it is necessary but not sufficient:

(6) a. I assessed candidates for the sales manager position. → Candidate assessment for the sales manager position.

b. I assigned John for the job. → *John assignment for the job

Furthermore, it is observed that semantic relation of the deverbal nominal to the verb is also a determiner of nominal syntactic structure. For example, nominals which incorporate verbal subject, object and indirect object roles cannot have verbal subject, object and indirect object arguments.

(7) He fights for the humankind. → The humankind fighter (Subject "he" cannot be realized in the nominal, no matter the syntactic position)

It seems that an intricate relationship between morphology, semantics and syntax is at play in producing a deverbal nominal, eluding theoretical approaches to most extent. This relationship is explored empirically with data mining and machine learning techniques. More specifically, it is investigated to see whether the seemingly unpredictable syntactic structure of deverbal nominals does not need as much degrees of freedom as the data itself, but self-organizes.

1.3 Organization of the Thesis

The rest of the thesis is divided into 6 chapters. In the next chapter follows NOMLEX is introduced. Then in the third chapter the problem is further analyzed, a hypothesis is formed and the roadmap for the solution is outlined. Related work is also summarized in this chapter. In the dataset chapter, data sources for the study is introduced and dataset creation steps are explained. In the fifth chapter division of the dataset into training and validation sets, models used in learning and performance evaluation metrics are explained. Results are discussed in the sixth chapter. The study is summarized in the last chapter and future work is suggested.

CHAPTER 2

NOMLEX: A LEXICON OF NOMINALIZATIONS

2.1 NOMLEX

NOMLEX (Macleod et al., 1998) is a lexicon of English deverbal nominals. Given the verb and the syntactic structure of the sentence, it specifies how subject, object, indirect object and oblique argument can be realized in a produced deverbal nominal. It consists of 1025 nominals derived from 941 verbs, which are found by searching for common nominalizer suffixes in the Wall Street Journal and Brown Corpus.

A sample NOMLEX entry is in the shape of a dictionary (Reeves et al., 1999), and contains the following:

- orthography of the deverbal nominal,
- the verb associated with the deverbal nominal,
- orthography of the plural form of deverbal nominal,
- ratio of the occurrences of the plural deverbal nominal to the occurrences of the singular deverbal nominal found in the British National Corpus,
- information of whether the deverbal nominal is only present in the plural form,
- aspect of the verb phrase incorporated by the deverbal nominal,
- information about existence and prevalence of a homograph noun,
- possible noun complements that can appear with the deverbal nominal,
- semantic attributes of the subject and object of the sentence the deverbal nominal is derived from,
- information of whether the subject or objects must be a plural or collective noun,
- information of whether object of the verb can also appear in subject position (alternation)
- information of whether the verb is in a passive or active form,

- list of sentence structures where the verb associated with the deverbal nominal can produce the deverbal nominal and the corresponding nominal positions where subject, direct and indirect object and oblique argument can be realized.

Deverbal nominal in NOMLEX are divided into 5 types according to which aspect of the verb phrase they incorporate (Reeves et al., 1999). These are:

- VERB-NOM: The deverbal nominal expresses the event or the state of the verb (e.g. destruction, abandonment).
- VERB-PART: The deverbal nominal incorporates verb particle (e.g. The takeover of the company by ITT $\rightarrow$ ITT took over the company).
- SUBJECT, OBJECT, IND-OBJECT: The deverbal nominal incorporates subject, object and indirect object role of the verb, respectively (e.g. teacher, appointee, payee).

Distribution of deverbal nominal types in the dataset is presented in Table 1.

Table 1: Types of deverbal nominals in NOMLEX.

Deverbal Nominal Type	n
VERB-NOM	859
VERB-PART	10
SUBJECT	106
OBJECT	19
IND-OBJECT	8

In NOMLEX, arguments of deverbal nominal are associated with semantic attributes (Reeves et al., 1999). The following attributes are applicable to both subject and direct object and also have an associated negative attribute with the prefix NOT :

- NUNIT: This attribute denotes units of measurement (e.g. foot, gallon, dollar).
- NHUMAN: This attribute denotes human entities (e.g. boy, cook, teacher). Names and pronouns are also denoted with this attribute.
- LOCATION: This attribute denotes proper place names (e.g. St George Church, Mount Magazine, IBM).
- NTIME: This attribute is more commonly used in the negative to indicate that a temporal element appearing in a potential subject or object position is not an argument.
- COMMUNICATOR: This attribute denotes superset of NHUMAN nouns and proper names, names of companies, etc.
- REFLEXIVE: This attribute is associated with an argument which in the sentence can only appear in reflexive form (e.g. behave, perjure).

There are 71 sentence structures in NOMLEX in which 941 verbs studied can produce at least one deverbal nominal. Distribution of sentence structures in the dataset is presented in Table 2. Examples for these sentence structures can be found in Appendix A.

Table 2: Sentence structures and corresponding number of verbs that produce at least one deverbal nominal when present in the sentence structure environment.

Sentence Structure	Number of Verbs
NOM-NP	805
NOM-NP-PP	437
NOM-INTRANS	321
NOM-PP	303
NOM-NP-AS-NP	141
NOM-PP-PP	57
NOM-THAT-S	48
NOM-S	43
NOM-INTRANS-RECIP	43
NOM-WH-S	43
NOM-NP-TO-INF-OC	39
NOM-POSSING	36
NOM-HOW-S	33
NOM-TO-INF-SC	30
NOM-P-POSSING	30
NOM-NP-TO-NP	30
NOM-NP-P-ING-OC	27
NOM-AS-NP	27
NOM-NP-P-POSSING	26
NOM-P-ING-SC	24
NOM-PP-THAT-S	24
NOM-P-WH-S	23
NOM-NP-ADVP	21
NOM-ING-SC	20
NOM-NP-PP-PP	19
NOM-ADVP	19
NOM-S-SUBJUNCT	15
NOM-FOR-TO-INF	13
NOM-PP-HOW-TO-INF	13
NOM-PART-PP	11
NOM-NP-ING-OC	10
NOM-EXTRAP-NP-S	10
NOM-NP-AS-ADJP	9
NOM-P-NP-ING	8
NOM-NP-S	8
NOM-PP-WH-S	8
NOM-PP-P-WH-S	7

Continued on next page

Table 2 – continued from previous page

Sentence Structure	Number of Verbs
NOM-NP-FOR-NP	7
NOM-NP-ING	7
NOM-NP-AS-ING	7
NOM-NP-AS-NP-SC	7
NOM-NP-P-WH-S	6
NOM-PART	6
NOM-NP-P-ING-SC	6
NOM-PART-NP-PP	6
NOM-PART-NP	6
NOM-NP-NUNITP-TO-RANGE	5
NOM-NP-NP	5
NOM-NP-AT-NUNITP-PRED	5
NOM-NUNITP-TO-RANGE	4
NOM-NP-ADJP-PRED	4
NOM-PP-P-POSSING	3
NOM-P-NP-TO-INF-VC	3
NOM-NP-TOBE	3
NOM-NP-P-NP-ING	3
NOM-POSSING-PP	3
NOM-NP-P-ING	3
NOM-ADVP-PP	2
NOM-P-NP-TO-INF-OC	2
NOM-NP-PP-AS-NP	2
NOM-PP-FOR-TO-INF	1
NOM-ADJP-PRED-RS	1
NOM-NP-TO-INF-SC	1
NOM-NUNITP-FROM-RANGE	1
NOM-P-NP-TO-INF	1
NOM-NP-WH-S	1
NOM-WORDS	1
NOM-PP-THAT-S-SUBJUNCT	1
NOM-PART-AS-NP	1
NOM-PART-NP-AS-NP	1
NOM-NP-NUNITP-FROM-RANGE	1

2.1.1 Realization of Subject, Direct Object and Indirect Object of the Sentence in a Deverbal Nominal

As previously noted, a NOMLEX entry includes list of sentence structures where the verb associated with the nominal can produce a deverbal nominal and syntactic positions where subject, direct object and indirect object can occur. These positions are:

(8) a. Possessive determiner

b. Prenoun noun modifier

c. Prepositional phrase

Usually, multiple deverbal nominals can be produced from the verb used in a given sentence structure and a verbal role can be realized in multiple positions across these deverbal nominals. A NOMLEX entry includes lists of valid positions for the verbal roles. However there are restrictions on how verbal roles can be realized in syntactic positions to create a deverbal nominal.

As a general rule, realization of verbal subject and object roles are optional, and it is explicitly stated in the entry if any of them must be realized. However, if present, indirect object role must be realized unless explicitly stated to be optional. If verbal roles are realized, they cannot be realized in multiple syntactic positions in a single nominal.

Only one role can fill possessive determiner position in a single nominal, and there cannot be multiple prepositional phrases with the same prepositional head. If there are multiple prepositional phrases they are assumed to occur in all orders. There can be multiple prenoun noun modifiers, and if this is the case they follow

$$subject > indirect\ object > object > oblique\ argument$$

order. Furthermore, a possessive determiner object cannot co-occur with a prenoun noun subject.

There can also be additional restrictions on how subject, object, indirect object and oblique argument can be realized that are on the verb and syntactic level and not covered by the above general rules and if this is the case these are also explicitly stated in the NOMLEX entry for the nominal. These additional restrictions can be in the form that a role cannot be realized in a certain position if another role has been realized in a specific position, or a role cannot be realized in a certain position if another role is absent in the deverbal nominal syntactic structure.

2.1.2 Realization of Oblique Argument of a Verb in a Deverbal Nominal

If present, oblique argument of a sentence must be realized in the syntactic structure of deverbal nominal, unless explicitly stated in the entry to be optional. Oblique arguments of verbs are usually realized as is in their deverbal nominal counterparts, however they can sometimes occur with additional or different prepositions. They can also appear in possessive determiner, prenoun noun modifier and prepositional phrase positions, like subject, object and indirect object roles. If they do, this information is also present in the NOMLEX entry for the deverbal nominal.

CHAPTER 3

NATURE OF THE PROBLEM AND RELATED WORK

3.1 The Lexicalist Hypothesis

The effort for relating deverbal nominals to their verbal counterparts and predicting their semantic, syntactic and morphological properties has branched into two approaches: the transformationalist approach and the lexicalist approach. The transformationalist approach uses syntactic derivation to relate deverbal nominals to their verbal counterparts; the lexicalist approach expresses the relation in the lexicon (Rozwadowska, 2017).

The transformationalist approach was first proposed by Lees (1962) and at the time had no alternative. Later, Chomsky (2014) proposed an extension of grammatical theory to incorporate syntactic features, which permitted a lexicalist approach to deverbal nominalizations. Then, Chomsky (1970) pointed out that unlike gerundive nominals, deverbal nominals cannot be explained by grammatical transformations because of the following reasons:

- Not every verb is able to produce a deverbal nominal
- The semantic relation between the verb and deverbal nominal varies
- Deverbal nominals have noun phrase internal structures

Chomsky (1970) continued to postulate that determiner information of nominalization lies in the lexicon and can be expressed as fixed selectional and strict subcategorizational features of lexical elements. As an example, Chomsky (1970) explains the following by hypothesizing that eager must be introduced into the lexicon with a strict subcategorization feature indicating that it can take a sentential complement, however easy must lack this feature since it cannot appear with a sentential complement, and the difference in features of "easy" and "eager" shows itself as one being able to produce a deverbal nominal and other can not:

(9) a. John is eager to please. → John's eagerness to please.

b. John is easy to please. → *John's easiness to please.

Chomsky (1970) has related nominalization to lexical elements' syntactic properties. Grimshaw (1990) takes a different approach from Chomsky (1970) and attributes nominalization to semantic properties. Grimshaw (1990) divides nouns into three types according to their semantic properties:

- Result nominals that express the result of the action denoted by the verb it is derived from.
- Simple event nominals that express the event of the verb but have no temporal organization.
- Complex event nominals that express the event of the verb, that are composed of two subevents and have temporal organization.

She then postulates that a nominal's ability to have grammatical arguments is dependent on its aspectual properties. She concludes only complex event nominals that have temporal organization can have arguments like verbs and therefore be considered as deverbal nominals in the sense used in the study, and possessives and by phrases that occur with result and simple event nominals have non-argument readings:

(10) a. Physics exam. (result nominal, no argument reading)

b. Boat race. (simple event nominal, no argument reading)

c. Recent Rome destruction. (aspectual modifier forces complex event reading, Rome has argument reading)

Similar to Grimshaw (1990), Anderson (1978) attributes nominalization to semantic properties of lexical items, and states arguments of verbs can map into prenoun noun modifiers of deverbal nominals if they are "changed, moved, altered in status or created". Building on Anderson (1978) and Grimshaw (1990), Doron & Rappaport-Hovav (1991) states the difference between verbs with affected objects and those with unaffected objects can be captured on the level of lexical representation, which they call event structure. Though it is demonstrated that the relation between argument realization and lexical properties identified by previous work is not as straightforward as assumed (Roeper, 1993) (Borer, 2003) (Newmeyer, 2009) (Alexiadou, 2011), it is believed that there is still merit in following a lexicalist approach.

The hypothesis of this work is that along with availability of a deverbal nominal, syntactic properties of resulting nominals are also determined by semantic and syntactic properties of lexical elements and although these properties have not been accurately identified by theoretical approaches, they can be learned by studying the relationships between lexical elements in large amounts of corpora.

3.2 Pre-trained Deep Learning Models

In recent years, research efforts have shifted from classical machine learning methods to deep learning techniques, mainly because they alleviated the feature engineering problem, and scarcity of computing power that made deep learning models unavailable was no longer an issue. This has led to significant advancements in NLP tasks across the board, mainly owing to pre-trained models.

Pre-trained models usually train unsupervised on large data such as Wikipedia, and learn common language representations which couldn't be achieved with the small sample data available for specific learning tasks (Qiu et al., 2020). This learned knowledge is then fine-tuned with the task data, and models trained with this method overperform models that only train on task data, mainly because latter models fail to generalize (Qiu et al., 2020) (Devlin et al., 2018) (He et al., 2020). In this light, this work fine-tunes pre-trained models to explore its hypothesis.

3.3 The Extended Research Question

941 verbs and 71 sentence structures studied in NOMLEX can produce 3025 unique deverbal nominal syntactic structures. This study tries to predict deverbal nominal syntactic structure using pre-trained deep learning models. In order to leverage the language models learned by pre-trained models, available sentences in various corpara having sentence structures and verbs studied in NOMLEX is intended to be used in fine-tuning.

3.4 Limitations of Unsupervised Pre-trained Deep Learning Models and Controlled Variables

Although unsupervised pre-trained deep learning models have proved themselves to be successful in various NLP tasks, they have their limitations. These limitations are discussed in the following sections and the scope of the research is narrowed down accordingly.

3.4.1 Semantic Relation Between Verb and Deverbal Nominal

It is observed that the semantic relation between the verb and deverbal nominal is a determiner of nominal syntactic structure, to give an example, incorporated role by the nominal is not realized in the nominal syntactic structure and it is expected that pre-trained models have some understanding of the semantic relation between the verb and deverbal nominal as part of the language model they have learned. However in order to make use of this information, sentence examples and orthography of the deverbal nominal must be linked in some way. It is not possible to feed features to a deep learning models like traditional machine learning algorithms and because of this reason semantic relationship between the verb and deverbal nominal is controlled. Deverbal nominals in NOMLEX are divided into 5 categories regarding how they semantically relate to their verbal counterparts and their distribution is presented in Table 1. VERB-NOM is chosen as the semantic relation to be studied because most number of verbs have this relation with their nominal counterparts, specifically 830.

3.4.2 Sentential Syntactic Structure

It is also observed that verbal syntactic environment is a determiner of nominal syntactic sturucture (Grimshaw, 1990) (Chomsky, 1970). However, it is unclear whether pre-trained models can learn syntactic structure from unsupervised data, though they can learn syntactic structure with supervised syntax-related pre-training tasks Sun et al. (2022). Subsequently, this study will focus on predicting syntactic structure of nominals derived from sentences with NOM-NP sentence structure, because most number of verbs nominalize when present in this sentence structure, specifically 805. NOM-NP sentence structure corresponds to sentences which have a subject, a transitive verb and an object, and example sentences for this structure can be found in Appendix A.

3.5 The Reformulated Research Question

There are 702 verbs present in NOMLEX that produce a deverbal nominal when present in NOM-NP sentence structure and have VERB-NOM semantic relation with their nominal counterparts and these verbs can produce 138 unique nominal syntactic structures. This shows that even when sentential syntactic environment and the relation between verb and nominal is controlled, there is still significant variability in deverbal nominal syntax. This study investigates whether this variability can be captured by the language models created by pre-trained models, if they are fine-tuned with available NOM-NP sentences of 702 verbs in the corpora.

3.5.1 Its Relation to Selectional Preferences

Since the semantic relation between verb and deverbal nominal and the sentence structure the deverbal nominal is derived from is controlled, and along with these, sentences available in corpora are used in fine-tuning, the only information provided for the pre-trained model to learn in the fine-tuning stage is which verbs are co-occurring with what arguments, or in other words, verb's selectional preferences. Roberts & Egg (2014) defines selectional preferences as follows:

> Selectional preferences Katz & Fodor (1963) Wilks (1975) Resnik (1993) are the tendency for a word to semantically select or constrain which other words may appear in a direct syntactic relation with it.

The following examples can be given to further demonstrate the concept:

(11) a. I ran a 10K marathon.

b. *My headphones ran a 10K marathon.

c. *I ran a spaghetti.

Although the above examples have the same syntactic structure, native speakers would attest that only the first example is well-formed. This is because headphones don't have the ability to run and spaghettis can't be run either.

Metheniti et al. (2020) finds that semantics of the predicate is an integral and influential factor for the selection of its arguments and selectional preferences are used successfully for various semantic tasks such as word-sense disambiguation (McCarthy & Carroll, 2003) and semantic role labeling (Gildea & Jurafsky, 2002). Since syntactic structures of deverbal nominals cannot be solely explained with variances in syntactic structures of sentences they are derived from, it is a possibility that syntactic structures of deverbal nominals can be influenced by selectional preferences of verbs they are derived from.

3.6 Choosing a Pre-trained Model

There are many different pre-trained models available, having different performances on different learning tasks because they differ in architecture, pre-training tasks and corpora. Naturally, a significant effort should be made to choose a model appropriate for the task at hand. It is important to further specify the nature of the problem, since it has a direct effect in pre-trained model choice.

Nominalization can be regarded as a generation problem, however, given the limited generative space, it can also be regarded as a classification problem. Since more than one deverbal nominal is produced per sentence most of the time, it is compelling to regard the nominalization problem as a multi-label classification problem.

3.6.1 Transformer Encoder Models

Vaswani et al. (2017) showed that a transformer model outperforms RNNs and LSTMs on translations tasks and attributed its success to its non-sequential nature: because it did not rely on past hidden states it did not suffer from long range dependency issues. Since then, transformer based models constitutes state-of-the-art and it seems natural to go for a transformer based model.

The original transformer in Vaswani et al. (2017) was a translation model and appropriately used an encoder–decoder architecture. The function of the encoding part of the architecture is to generate contextual word encodings. Then decoding part uses these encodings to generate an output sequence. Transformer models meant for natural language understanding tasks such as text classification and question answering only need the encoding part of the architecture to learn the contextual word encodings and fine-tune a task specific last layer to make use of these encodings (Devlin et al., 2018). Subsequently, this work will recruit various transformer encoder models with different architectures and pre-training tasks to explore its hypothesis.

3.7 Related Work

Lapata (2002) focuses on the disambiguation of prenoun noun modifiers of deverbal nominals who realize either the subject or the object in that position. They define the problem as a binary classification task, make the assumption that the relation of the nominalized head and its modifier noun is the same as the relation between the latter and the verb from which the head is derived and use a log-likelihood model for prediction. They conclude that a combination of smoothing methods for data sparseness and introduction of shallow contextual information achieves 86.1% accuracy, compared to 61.5% baseline.

Grover et al. (2005) focuses on the disambiguation of prenoun noun modifiers of deverbal nominals similar to Lapata (2002) however they also take into account prepositional objects, and treat the problem as a 15 class classification task. Following Lapata (2002) the argument relation between a deverbal head and its modifier is approximated by the relation of the underlying verb and its arguments. They use a decision tree learner and unlike Lapata (2002) use verb-argument counts as features. They report 66.9% accuracy with the best performing feature set, compared to 46.65% baseline.

Similar to Lapata (2002) and Grover et al. (2005), Nicholson & Baldwin (2005) focuses on the disambiguation of noun modifiers of deverbal nominalizations. They treat the problem as a 3-class classification task and classify the modifier as subject, object or the prepositional object of the verb the deverbal nominal is derived from. They consider each occurrence of a verb-noun pair to be a normally distributed binomial trial for two relations under consideration, and calculate corresponding z-scores. They attribute the relation that have the highest z-score to the modifier. They achieve 70% accuracy in classifying the relationship as subject or object and 57% accuracy in classifying the relationship as subject, object or prepositional object.

Gurevich & Waterman (2009) focuses of mapping arguments of deverbal nouns to verbs they are derived from. They restrict their scope to deverbal nouns that are derived from transitive verbs that realize only one argument in the possessive modifier position or the prepositional phrase headed by of. They hypothesize that arguments that are mapped into subject and object roles by verbs will also map into the same roles by deverbal nominals. They test their hypothesis on three models and conclude that the model including lexical information of arguments performs best with an F_1 score of 0.85 when predicting the role in the possessive determiner position, however the model that uses general preferences performs best when predicting the role in the of prepositional phrase with F_1 score of 0.82.

CHAPTER 4

DATASET

The effort for finding available NOM-NP sentences of 702 verbs that nominalize in this sentence structure and have VERB-NOM semantic relation with their deverbal nominal counterparts is described in the following sections.

4.1 Data Sources

The sentences in the dataset are collected from two data sources: Gigaword (Graff et al., 2003) and Wikipedia (Ortman, 2018).

4.1.1 Gigaword

English Gigaword (Graff et al., 2003) is a comprehensive archive of newswire text data that has been acquired over several years by the Linguistic Data Consortiume. The seven English newswire sources included in Gigaword are:

- Agence France-Presse, English Service
- Associated Press Worldstream, English Service
- Central News Agency of Taiwan, English Service
- Los Angeles Times/Washington Post Newswire Service
- Washington Post/Bloomberg Newswire Service
- New York Times Newswire Service
- Xinhua News Agency, English Service

The dataset used in this study is the annotated and preprocessed with Stanford CoreNLP tools (Manning et al., 2014) version by Rush et al. (2015), as this is the only publicly available version. This dataset was meant to be used in learning of sentence summarization tasks and consists of sentence-article pairs, however only article portions of the dataset is used in this study.

4.1.2 Wikipedia

This is a collection of 7.8 million sentences retrieved from August 2018 English Wikipedia (Ortman, 2018). This collection is formed by parsing the text with the SpaCy library (Honnibal & Montani, 2017) and filtering out the sentences that require citations, have unmatched parenthesis or brackets and are shorter than 3 or longer than 255 letters. Duplicate sentences are also removed from the collection.

4.2 Parsing and Filtering Sentences

There are 11.8 million sentences combined in the aforementioned two data sources. These sentences are first filtered by searching for base and conjugated forms of 702 verbs in NOMLEX using mlconjug3 library (Diao, 2021).

4.2.1 Parsing Sentences Using Combinatory Categorical Grammar

Steedman & Baldridge (2011) defines Combinatory Categorial Grammar (CCG) as follows:

> Combinatory Categorial Grammar (CCG) is a radically lexicalized theory of grammar in which all language-specific information, including the linear order of heads, arguments, and adjuncts, is specified in the lexicon, from which it is projected onto sentences by language-independent universal type dependent combinatory rules of low "slightly non-context-free" expressive power, applying to strictly adjacent phonologically-realised categories. Syntactic and phonological derivation are isomorphic, and are synchronously coupled with semantic composition in a purely type-dependent rule-to-rule relation.

In CCG framework, because all language specific information is specified in the lexicon, words are assigned to different categories depending on the sentence structure they are present in, meaning each word's category contains some information about the sentence structure, especially the verb. This makes filtering data for sentence structure especially efficient. The following are examples of CCG parse trees in the form of nested lists for two different sentence structures sharing the same verbs:

(12) a. I warned him. → (<T S[dcl] 1 2> (<L NP POS POS I NP>) (<T S[dcl]\NP 0 2> **(<L (S[dcl]\NP)/NP POS POS warned (S[dcl]\NP)/NP>)** (<L NP POS POS him NP>)))

b. I warned him about her. → (<T S[dcl] 1 2> (<L NP POS POS I NP>) (<T S[dcl]\NP 0 2> (<T (S[dcl]\NP)/PP 0 2> **(<L ((S[dcl]\NP)/PP)/NP POS POS warned ((S[dcl]\NP)/PP)/NP>)** (<L NP POS POS him NP>)) (<T PP 0 2> (<L PP/NP POS POS about PP/NP>) (<L NP POS POS her NP>))))

Checking the category of the verb, which can be found in limited places in the parse tree because of fixed word order in English, is enough to understand the sentence structure of examples above.

Because of the advantages of using CCG parses for filtering data, sentences filtered by conjugated forms of verbs in NOMLEX are then parsed by EasyCCG CCG parser (Lewis & Steedman, 2014), the most successful of its kind to date.

After the sentences are parsed with EasyCCG, first, parse trees of sentences that might have NOM-NP sentence structure are extracted. This corresponds to subtrees that diverge from a S[dcl] category, which amounts to a well formed sentence, and does not contain another S[dcl] category node. This means a sentence that is part of another sentence, for example a sentence part of a wh-sentence clause, is extracted and sentences that contain sentences, like sentences that contain that-sentence clauses, are discarded because they can never have NOM-NP sentence structure. Then verb and verb complement of these sentences are further analyzed.

4.2.1.1 Checking the Verb

The sentences in the data sources were filtered by checking for base and conjugated forms of verbs in NOMLEX that both nominalize in NOM-NP structure and have VERB-NOM semantic relation with their nominal counterparts. However, further checking needs to be done to make sure these verbs are part of the verb of the sentence and not for example part of a gerund complement.

In a CCG parse tree of a sentence, because of fixed word order of English, left child always includes the subject and has NP category, and the right child has S\NP category and consists of verb and its complement, if exists. If there are no auxillary verbs or verb modifiers in the sentence, verb of the sentence is always found in the left child of the first right child, however if auxillary verbs or verb modifiers exist, first auxiallary verb is the left child and the primary verb is in the right child or the verb modifier is the first child and the verb is in the second child. Because of this variability, primary verb and possible auxillary verbs and verb modifiers of the sentence are extracted from the right child of the parse by looking for the first node with NP category, which stands for noun phrase. Words up to NP category node is considered the verb of the sentence to be further checked and rest of the words are considered to be the verb complement. If the right child of the parse does not contain a node with NP category, for example if the verb is used in a intransitive meaning, the sentence is discarded as not having NOM-NP structure.

After the verb's extraction from the parse, primary verb, which is the last word in the extracted verb structure, is checked to be one of the base or conjugated forms of the 702 verbs that nominalize in NOM-NP sentence structure and have VERB-NOM semantic relation with their nominal counterparts. Then the primary verb is checked to have S\NP/NP category. Verbs that are used in sentences that only have a subject and an object have this CCG category, and these sentences correspond to the NOM-NP sentence structure in NOMLEX. If auxillary verbs exist in the verb structure, they are checked to have the structure of one of the English tenses or modals.

4.2.1.2 Checking the Verb Complement

Since verb of the sentence is checked to have S\NP/NP category, it is known that verb complement includes a noun phrase. However, further checking is necessary to rule out sentence structures that might not have NOM-NP sentence structure.

It is rare to find sentences with only a subject and an object in corpora. Most sentences have adjuncts, are compounded with another sentence, or have compound objects. To increase dataset size, these sentences are not discarded completely. Adjuncts are removed from the sentences and only the first noun of noun phrases that are conjoined by comma, "and" or "but" are left in the sentence. Latter part of compound sentences are also removed. Adjectives that appear after the object are also removed because this sentence structure is recognized by NOMLEX as another type of sentence.

4.3 Classes

The classes are created using the information in NOMLEX (Macleod et al., 1998) and complete set of classes with their frequencies in the dataset of 702 verbs is presented in Table 3.

Table 3: Deverbal nominal syntactic structures and corresponding number of verbs that produce a VERB-NOM deverbal nominal with the syntactic structure when present in NOM-NP sentence structure.

Deverbal Nominal Syntactic Structure	Number of Verbs
The NOUN by SUBJECT	653
The NOUN of OBJECT	630
SUBJECT's NOUN of OBJECT	618
The NOUN by SUBJECT of OBJECT	607
SUBJECT's NOUN	584
OBJECT's NOUN	437
SUBJECT's OBJECT NOUN	422
The OBJECT NOUN	419
The OBJECT NOUN by SUBJECT	405
The SUBJECT NOUN	335
The SUBJECT NOUN of OBJECT	305
The SUBJECT OBJECT NOUN	232
The NOUN of SUBJECT	143
The OBJECT NOUN of SUBJECT	68
OBJECT's NOUN of SUBJECT	67
The NOUN to OBJECT	40
SUBJECT's NOUN to OBJECT	37
The NOUN by SUBJECT to OBJECT	33
The NOUN from SUBJECT	20
The SUBJECT NOUN to OBJECT	18
OBJECT's NOUN with SUBJECT	16
The NOUN with SUBJECT	16
The NOUN on OBJECT	16
The NOUN of SUBJECT to OBJECT	15
The NOUN from SUBJECT of OBJECT	15
SUBJECT's NOUN for OBJECT	14

Continued on next page

Table 3 – continued from previous page

Deverbal Nominal Syntactic Structure	Number of Verbs
The NOUN for OBJECT	14
SUBJECT's NOUN on OBJECT	14
The NOUN by SUBJECT on OBJECT	14
The SUBJECT NOUN on OBJECT	13
The OBJECT NOUN from SUBJECT	13
The NOUN of SUBJECT for OBJECT	11
The NOUN by SUBJECT for OBJECT	11
OBJECT's NOUN over SUBJECT	11
The NOUN over SUBJECT	11
The NOUN with SUBJECT of OBJECT	11
The NOUN of SUBJECT on OBJECT	10
The NOUN over SUBJECT of OBJECT	9
OBJECT's NOUN at SUBJECT	8
The NOUN at SUBJECT	8
The NOUN from SUBJECT to OBJECT	8
OBJECT's NOUN from SUBJECT	8
SUBJECT's NOUN against OBJECT	7
The NOUN by SUBJECT against OBJECT	7
The NOUN against OBJECT	7
The OBJECT NOUN with SUBJECT	7
The SUBJECT NOUN against OBJECT	6
The SUBJECT NOUN for OBJECT	6
The NOUN at SUBJECT of OBJECT	5
OBJECT's NOUN in SUBJECT	5
The NOUN in SUBJECT of OBJECT	5
The NOUN in SUBJECT	5
The NOUN of SUBJECT against OBJECT	5
The NOUN between SUBJECT	5
The NOUN between SUBJECT of OBJECT	4
The NOUN among SUBJECT	4
The NOUN by SUBJECT in OBJECT	4
The NOUN in OBJECT	4
The OBJECT NOUN at SUBJECT	3
The OBJECT NOUN over SUBJECT	3
The NOUN with SUBJECT to OBJECT	3
The NOUN of SUBJECT at OBJECT	3
The SUBJECT NOUN at OBJECT	3
SUBJECT's NOUN at OBJECT	3
The NOUN by SUBJECT at OBJECT	3
The NOUN at OBJECT	3
OBJECT's NOUN between SUBJECT	3
The OBJECT NOUN between SUBJECT	3
The NOUN among SUBJECT of OBJECT	3
The NOUN by SUBJECT with OBJECT	3

Continued on next page

Table 3 – continued from previous page

Deverbal Nominal Syntactic Structure	Number of Verbs
The NOUN with OBJECT	3
SUBJECT's NOUN about OBJECT	3
The NOUN by SUBJECT about OBJECT	3
The NOUN about OBJECT	3
SUBJECT's NOUN in OBJECT	3
The SUBJECT NOUN in OBJECT	3
The NOUN at SUBJECT to OBJECT	2
The NOUN over SUBJECT to OBJECT	2
OBJECT's NOUN to SUBJECT	2
The NOUN to SUBJECT of OBJECT	2
The NOUN to SUBJECT	2
OBJECT's NOUN among SUBJECT	2
The OBJECT NOUN among SUBJECT	2
SUBJECT's NOUN with OBJECT	2
The NOUN of SUBJECT with OBJECT	2
The NOUN from SUBJECT on OBJECT	2
The NOUN of SUBJECT about OBJECT	2
The NOUN of SUBJECT over OBJECT	2
The SUBJECT NOUN over OBJECT	2
SUBJECT's NOUN over OBJECT	2
The NOUN by SUBJECT over OBJECT	2
The NOUN over OBJECT	2
The NOUN amongst SUBJECT	2
The NOUN by SUBJECT among OBJECT	2
The NOUN among OBJECT	2
SUBJECT's NOUN into OBJECT	2
The NOUN by SUBJECT into OBJECT	2
The NOUN into OBJECT	2
The NOUN from SUBJECT for OBJECT	2
The NOUN of SUBJECT in OBJECT	2
SUBJECT's NOUN upon OBJECT	2
The NOUN by SUBJECT upon OBJECT	2
The NOUN upon OBJECT	2
OBJECT's NOUN for SUBJECT	1
The NOUN for SUBJECT of OBJECT	1
The NOUN for SUBJECT	1
The SUBJECT NOUN with OBJECT	1
The NOUN amongst SUBJECT about OBJECT	1
The NOUN amongst SUBJECT on OBJECT	1
The NOUN amongst SUBJECT over OBJECT	1
The NOUN between SUBJECT about OBJECT	1
The NOUN between SUBJECT on OBJECT	1
The NOUN between SUBJECT over OBJECT	1
The NOUN among SUBJECT about OBJECT	1

Continued on next page

Table 3 – continued from previous page

Deverbal Nominal Syntactic Structure	Number of Verbs
The NOUN among SUBJECT on OBJECT	1
The NOUN among SUBJECT over OBJECT	1
The SUBJECT NOUN about OBJECT	1
OBJECT’s NOUN about SUBJECT	1
The NOUN about SUBJECT of OBJECT	1
The NOUN about SUBJECT	1
SUBJECT’s NOUN among OBJECT	1
The NOUN over SUBJECT among OBJECT	1
OBJECT’s NOUN amongst SUBJECT	1
The OBJECT NOUN amongst SUBJECT	1
The NOUN amongst SUBJECT of OBJECT	1
The NOUN with SUBJECT among OBJECT	1
The NOUN in SUBJECT for OBJECT	1
The NOUN in SUBJECT to OBJECT	1
The NOUN with SUBJECT for OBJECT	1
The OBJECT NOUN to SUBJECT	1
The SUBJECT NOUN upon OBJECT	1
SUBJECT’s NOUN towards OBJECT	1
SUBJECT’s NOUN toward OBJECT	1
The NOUN towards OBJECT	1
The NOUN toward OBJECT	1
The NOUN of SUBJECT into OBJECT	1
The SUBJECT NOUN into OBJECT	1
The NOUN from SUBJECT against OBJECT	1

CHAPTER 5

LEARNING

5.1 Splitting the Dataset into Training and Validation Sets

When initial tests are done on training and validation sets where sentences including the same verb existed on both datasets, the models had 99.9% performance on all of the evaluation metrics. This means that pre-trained models are able to locate the verb of a sentence and make the relation that verb of the sentence determines deverbal nominals produced when sentence structure is controlled. To make things more interesting, verbs are split into training and test sets, meaning learning is done on the dataset which sentence examples of a particular verb is only in the training set or the validation set.

5.1.1 Addressing Data Sparseness in the Dataset

As seen in Table 3, although 138 unique deverbal nominal syntactic structures are produced by 702 verbs, most of these are produced by three or less verbs and only 37 of them are produced by at least 10 verbs. Because it wouldn't be reasonable to learn classes that are only produced by few verbs, only the most represented classes are included in the dataset. It is observed in Table 3 that classes that include only the prepositional phrases headed by "of" and "by" are significantly produced, therefore classes including only these prepositional phrases are included in the dataset.

Futhermore, number of NOM-NP sentences found in the data sources varied drastically for the 702 verbs, ranging from 0 to 85.931. Verbs without sufficient number of examples, chosen to be 1000, are not included in the dataset because it is supposed that the model will not be able to generalize for these verbs. Furthermore, to avoid over representation of verbs with more than 5000 examples, only 5000 of the sentences are selected at random.

The above modifications reduced the number of verbs in the dataset to 131, and number of deverbal nominal classes to 15. Distribution of deverbal nominal classes in the final dataset is presented in Table 4. Number of sentences in the dataset is 362.387, and average number of sentences per verb is 2766.

5.1.2 Stratification

Using k-fold cross validation when training and evaluating models gives a better measurement of model performance because the model is trained and tested in all of the data samples. Model performance is averaged across folds, minimizing the effect of a possible "easy" or "hard" validation set.

Table 4: Deverbal nominal syntactic structures that are included in the final dataset and percentage of verbs that produce a deverbal nominal with the syntactic structure.

Deverbal Nominal Syntactic Structure	Verbs (%)
The NOUN by SUBJECT	95
SUBJECT's NOUN	88
SUBJECT's NOUN of OBJECT	87
The NOUN by SUBJECT of OBJECT	86
The NOUN of OBJECT	85
SUBJECT's OBJECT NOUN	83
The OBJECT NOUN by SUBJECT	63
The OBJECT NOUN	59
OBJECT's NOUN	59
The SUBJECT NOUN	54
The SUBJECT NOUN of OBJECT	48
The SUBJECT OBJECT NOUN	38
The NOUN of SUBJECT	24
OBJECT's NOUN of SUBJECT	11
The OBJECT NOUN of SUBJECT	11

Empirical analysis has shown that the best results are attained when 80% of data is included in the training set and 20% in the validation set, and a pedagogical explanation is given in Gholamy et al. (2018). Subsequently, the dataset is divided into 5 folds and every fold is used as a validation set, rest of the folds forming the training set. The overall model performance is found by averaging the performance of the model across folds.

As seen in Table 4 prevalence of deverbal nominals syntactic structures produced by verbs in the dataset varies greatly, with most prevalent one being produced by 95% of the verbs in the dataset and least prevalent ones being produced by only 11%. Because some of the syntactic structures are produced by very few verbs, there is a possibility that if random sampling is done these classes might not be present in some folds. Stratified sampling is used to ensure all of the classes are represented across folds, which is defined by Sechidis et al. (2011) as follows:

> Stratified sampling is a sampling method that takes into account the existence of disjoint groups within a population and produces samples where the proportion of these groups is maintained.

Iterative stratification algorithm for multi-label stratification (Sechidis et al., 2011) (Szymański & Kajdanowicz, 2017) implemented by scikit-multilearn library (Szymański & Kajdanowicz, 2017) is used to create the folds. Distribution of classes across folds are presented in Table 5.

Table 5: Distribution of deverbal nominal syntactic structures across folds.

Deverbal Nominal Syntactic Structure	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
The NOUN by SUBJECT	94	95	96	96	95
SUBJECT'S NOUN	89	88	90	88	88
SUBJECT'S NOUN of OBJECT	88	88	89	87	87
The NOUN by SUBJECT of OBJECT	86	87	87	86	86
The NOUN of OBJECT	87	87	84	85	85
SUBJECT'S OBJECT NOUN	63	63	65	64	62
The OBJECT NOUN by SUBJECT	59	60	62	61	59
The OBJECT NOUN	60	59	60	60	58
OBJECT'S NOUN	58	57	62	58	63
The SUBJECT NOUN	50	56	56	55	55
The SUBJECT NOUN of OBJECT	43	50	51	49	47
The SUBJECT OBJECT NOUN	36	38	39	41	37
The NOUN of SUBJECT	27	22	24	27	22
OBJECT'S NOUN of SUBJECT	11	12	12	13	9
The OBJECT NOUN of SUBJECT	12	12	12	13	8

5.2 Evaluation Techniques

Unlike single-label classification where inference results can be considered true or false, in multi-label classification, predicted results can be completely correct, partially correct or completely incorrect. Because of this, multi-label classification requires different metrics than those used in single-label classification (Tsoumakas & Katakis, 2007) (Sorower, 2010).

Also unlike single-label classification, there are two ways to aggregate metrics in multi-label classification. First is sample-based aggregation where the metric is first computed on the example and scores of examples are averaged. The second is label-based aggregation where metrics are evaluated on label level and averaging is done for the scores of label classes (Sorower, 2010). In this study, models will be evaluated in both ways because both ways provide different insights to the model's performance. While sample based evaluation is useful for assessing overall performance of the model, label based evaluation techniques provide more insight as to how models perform on different labels.

5.2.1 Sample-based Evaluation Techniques

Let L be the finite set of class labels and D be the multi-label evaluation data set, consisting of $|D|$ multi-label examples $(x_i, Y_i), i = 1..|D|, Y_i \subseteq L$, H be the multi-label classifier and $Z_i = H(x_i)$ be the set of labels predicted by H for example x_i.

5.2.1.1 Accuracy

Sorower (2010) defines multi-label accuracy for each sample as the proportion of the predicted correct labels to the total number of predicted and actual labels for that sample. Overall accuracy is the average across all samples and calculated as follows:

$$Accuracy = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \tag{1}$$

5.2.1.2 Hamming Loss

Schapire & Singer (2000) and Tsoumakas & Katakis (2007) consider hamming loss as a performance metric for multi-label classification tasks, that is defined as:

$$HammingLoss = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \Delta Z_i|}{|L|} \tag{2}$$

Δ stands for the symmetric difference of two sets which corresponds to the XOR operation in Boolean logic. According to Sorower (2010), hamming loss takes into account both the prediction error and the missing error. Scikit-learn library's (Pedregosa et al., 2011) implementation of Hamming Loss algorithm is used in the study.

5.2.1.3 Precision

Precision is the proportion of predicted correct labels to the total number of predicted labels. Sample based precision is the average across all samples and calculated as follows:

$$Precision_s = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Z_i|} \tag{3}$$

5.2.1.4 Recall

Recall is the proportion of predicted correct labels to the total number of actual labels. Sample based recall is the average across all samples and calculated as follows:

$$Recall_s = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Y_i|} \tag{4}$$

5.2.1.5 F_1 Score

Sample F_1 score is the harmonic mean of precision and recall that is averaged over samples:

$$F_1Score_s = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{2|Y_i \cap Z_i|}{|Y_i| + |Z_i|} \tag{5}$$

5.2.2 Label-based Evaluation Techniques

Precision, recall and F_1 score can also be calculated for each label and then be averaged over all labels. In addition to these, area under precision vs. recall and ROC curves can be used as label-based evaluation metrics.

5.2.2.1 Area Under ROC Curve

A receiver operating characteristic curve, or ROC curve, is the graph of true positive rate vs false positive rate on various threshold settings. For a random model, the graph would approximate to $x = y$ line, and for a model with good predictive power, the graph would lie on the upper side of an imaginary $x = y$ line, because true positive rate would always be higher than false positive rate at any threshold value. Because of the nature of the graph, area under the ROC curve can be used as a performance metric. An advantage of using area under ROC curve is that it results in a performance metric independent of threshold value chosen for the classification, enabling more accurate interpretation of model performance. However, Saito & Rehmsmeier (2015) finds from empirical evidence that false positive rates are reduced in imbalanced classes because of high numbers of true negatives and ROC curves should not be used to interpret model performance on imbalanced classes.

5.2.2.2 Area Under Precision vs. Recall Curve

Precision vs. Recall curve is calculated by computing precision and recall in various threshold settings. For a model with good predictive power, the graph would approximate to $y = 1$ line, however in most cases precision decreases as recall increases. Because of the nature of the graph, area under precision vs. recall curve can be used as a performance metric and similar to area under ROC curve, it has the advantage of being independent of threshold value chosen for classification. Furthermore, Saito & Rehmsmeier (2015) concludes that area under precision vs. recall curve can be used to interpret model performance on imbalanced classes, which is not the case for area under ROC curve.

5.3 Models

Experiments with different models with different architectures and pre-training models are done to determine the best model for the research question. For experiments, current state-of-the-art models are chosen.

5.3.1 BERT

BERT, standing for bidirectional encoder representations from transformers, is a transformer-encoder architecture based pre-trained model (Devlin et al., 2018). Unlike its predecessors, it is trained on next sentence prediction and masked language modeling, which enables the model to "see" both left and the right context when making predictions. BERT needs training of only one additional layer to create models for various NLP tasks and shows fine-tuned pre-trained models reduce the need for task specific architectures by advancing state of the art in both various sentence-level and token-level tasks.

5.3.2 XLNet

Yang et al. (2019) finds that artificial symbols like [MASK] used by BERT for masked language modeling to facilitate bidirectional learning are absent during fine-tuning, resulting in a pretrain-finetune discrepancy. Furthermore, they report that since predicted tokens are masked in the input, BERT is unable to model joint probability, causing BERT to assume that predicted tokens are independent of each other while long range dependencies are prevalent in natural language. To overcome BERT's shortcomings, Yang et al. (2019) proposes XLNet based on a novel generalized permutation language modeling objective, which allows modeling of joint probability and removes the pretrain-finetune discrepancy present in BERT because training data is not corrupted. XLNet also integrates ideas from Transformer-XL (Dai et al., 2019) for its architecture. Authors report that XLNet significantly outperforms BERT on various benchmarks.

5.3.3 RoBERTa

Liu et al. (2019) finds that the BERT model is significantly undertrained and reports hyperparameter tuning and design changes that improves BERT substantially. They:

- train the model longer, with bigger batches, over more data.
- only train on masked language modeling.
- train on longer sequences, providing more context to model.
- dynamically change the masking pattern applied to the training data.

They state that with these improvements, BERT is able to match the performance of every model that is released after itself. They call their model RoBERTa for robustly optimized BERT approach.

5.3.4 DeBERTa

He et al. (2020) introduces DeBERTa: Decoding-enhanced BERT with disentangled attention that improves the BERT and RoBERTa models using two novel techniques. They introduce disentangled attention mechanism, where each word is represented by not one but two vectors: one for encoding their position and one for encoding their contents. They also use an enhanced mask decoder, enabling incorporation of absolute positions of words in the decoding layer to predict the masked tokens. They compare DeBERTa to RoBERTa and report that a DeBERTa model trained on half of the data outperforms the latter on various NLP tasks. They also report that DeBERTa surpasses human performance on SuperGLUE benchmark (Wang et al., 2019) for the first time in terms of macro-average score.

5.4 Hugging Face Transformers

For the training of previously mentioned models, Hugging Face Transformers library is used, which is defined by Wolf et al. (2020) as follows:

> Transformers is a library dedicated to supporting Transformer-based architectures and facilitating the distribution of pretrained models.

Transformers library hosts a wide range of transformer based models, provides tools for fine-tuning models for various learning tasks and facilitates effortless switching between models.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 Results

6.1.1 Best Model

Comparison of models' performance on sample-based and label-based performance techniques are presented in Table 6 and Table 7.

In sample based performance metrics RoBERTa and DeBERTa models perform similarly on F_1 score, however RoBERTa model has higher precision and lower recall values, meaning compared to the DeBERTa model it is more sensitive but less specific. In terms of accuracy DeBERTa model outperforms RoBERTa model by a small margin, and in terms of hamming loss DeBERTa model outperforms RoBERTa model by a respectable margin.

In label based performance metrics, RoBERTa model outperforms DeBERTa model on F_1 score, recall and PR AUC metrics, nonetheless DeBERTa model outperforms RoBERTa on precision and ROC AUC. This again indicates that RoBERTa model is more sensitive but less specific compared to the DeBERTa model.

Label-based performance metrics are lower than sample-based performance metrics where they are comparable, which can be explained by the fact that since label-based performance metrics are averaged over labels, they give more weight to labels that are less represented in the dataset compared to sample-based performance metrics. It is a possibility that models fail to generalize or overgeneralize on less prevalent labels because of lack of representation. It is worthwhile to mention that RoBERTa model outperforms DeBERTa model in label-based F_1 score, distinguishing its ability to generalize with limited samples.

6.1.2 Model Performances on Classes

Model's performances on classes are presented in Table 8, 9, 10 and 11. Precision vs. recall and ROC curves of models are also presented in Appendix B.

A baseline F_1 score for evaluating results can be calculated for classes by assuming the model predicts all of the classes for every sentence. The resulting recall scores would be 1 for every class, and

Table 6: Comparison of models' performance on sample-based performance metrics.

Performance Metric	BERT	RoBERTa	XLNet	DeBERTa
Precision	0.798	0.756	0.785	**0.800**
Recall	0.788	**0.856**	0.789	0.814
F_1 Score	0.760	**0.779**	0.758	**0.779**
Accuracy	0.641	0.663	0.634	**0.664**
Hamming Loss	0.274	0.263	0.271	**0.247**

Table 7: Comparison of models' performance on label-based performance metrics.

Performance Metric	BERT	RoBERTa	XLNet	DeBERTa
Precision	0.757	0.760	0.765	**0.791**
Recall	0.676	**0.785**	0.687	0.724
F_1 Score	0.674	**0.750**	0.677	0.725
ROC AUC	0.607	0.685	0.686	**0.712**
PR AUC	0.727	**0.803**	0.759	0.773

precision would be equal to the prevalence of the class in the dataset. Baseline F_1 scores of classes and highest F_1 scores are presented in Table 12

It it seen in Table 12 that for 10 of the 15 classes at least one model performs better than the baseline on F_1 score. When the remaining five classes are studied, it is seen that two of the three classes that realize the subject in possessive modifier position are unsuccessfully learned. This observation also applies to the classes that realize the subject in the prenoun noun modifier position. Furthermore it is observed that classes that only realize the subject or the object in the possessive determiner position are also unsuccessfully learned.

6.2 Discussion

When a preliminary study was conducted with training and validation sets where sentences including the same verb existed on both datasets, the models had 99.9% performance on all of the evaluation metrics. This means that pre-trained models are able to locate the verb of a sentence and make the relation that verb of the sentence determines deverbal nominals produced when sentence structure is controlled.

Because sentence examples of a particular verb is included in the validation set or the training set, the validation set consisted of sentence examples of verbs that are never seen by the model in the training phase. By design, the study assumes that verbs are organized among themselves according deverbal syntactic structures they produce. Thereby, any regularities concluded from study would also confirm this organization.

Table 8: BERT model's performance on classes.

Class	Precision	Recall	F1 Score	PR AUC	ROC AUC
The NOUN by SUBJECT	0.977	0.999	0.988	0.956	0.162
SUBJECT's NOUN	0.873	0.981	0.924	0.918	0.711
SUBJECT's NOUN of OBJECT	0.931	0.963	0.947	0.976	0.867
The NOUN by SUBJECT of OBJECT	0.916	0.934	0.927	0.956	0.774
The NOUN of OBJECT	0.860	0.943	0.900	0.814	0.545
SUBJECT's OBJECT NOUN	0.741	0.785	0.762	0.668	0.566
The OBJECT NOUN by SUBJECT	0.677	0.695	0.686	0.628	0.530
The OBJECT NOUN	0.693	0.741	0.717	0.630	0.525
OBJECT's NOUN	0.630	0.675	0.652	0.805	0.674
The SUBJECT NOUN	0.600	0.716	0.652	0.672	0.539
The SUBJECT NOUN of OBJECT	0.527	0.641	0.578	0.646	0.575
The SUBJECT OBJECT NOUN	0.553	0.500	0.525	0.623	0.574
The NOUN of SUBJECT	0.838	0.242	0.375	0.600	0.824
OBJECT's NOUN of SUBJECT	0.582	0.009	0.018	0.389	0.624
The OBJECT NOUN of SUBJECT	0.954	0.308	0.466	0.613	0.814

When the models are evaluated on classes, it is seen that model's performances varies greatly. For 10 out of 15 classes, at least one model performs better than baseline. However some of these improvements are more statistically significant than others. Three patterns can be observed in unsuccessfully learned classes and the observations suggest that determiners of an argument's occurrence in a syntactic position is dependent on both the argument and the syntactic position in question along with realization of other arguments in the deverbal nominal. In the light of these, it is concluded that determiners of an argument's realization in a syntactic position is dependent on both the argument and the syntactic position in question along with realization of other arguments. The presented method was able to learn these determiners to various extents.

It should also be noted that the experiments are done on base models rather than large models with more parameters because of computing power limitations and only 131 of 702 eligible verbs are included in the dataset because of lack of examples for less available verbs. The outcomes of the study can be improved upon eliminating these technical limitations.

Table 9: XLNet model's performance on classes.

Class	Precision	Recall	F1 Score	PR AUC	ROC AUC
The NOUN by SUBJECT	0.977	0.996	0.987	0.996	0.840
SUBJECT's NOUN	0.872	0.972	0.919	0.918	0.835
SUBJECT's NOUN of OBJECT	0.948	0.964	0.956	0.956	0.819
The NOUN by SUBJECT of OBJECT	0.925	0.956	0.940	0.968	0.843
The NOUN of OBJECT	0.881	0.947	0.913	0.885	0.670
SUBJECT's OBJECT NOUN	0.712	0.792	0.749	0.716	0.604
The OBJECT NOUN by SUBJECT	0.685	0.777	0.728	0.673	0.602
The OBJECT NOUN	0.667	0.749	0.705	0.619	0.529
OBJECT's NOUN	0.644	0.711	0.676	0.769	0.644
The SUBJECT NOUN	0.576	0.661	0.616	0.650	0.497
The SUBJECT NOUN of OBJECT	0.525	0.574	0.549	0.587	0.552
The SUBJECT OBJECT NOUN	0.585	0.554	0.569	0.616	0.563
The NOUN of SUBJECT	0.789	0.236	0.364	0.729	0.773
OBJECT's NOUN of SUBJECT	0.833	0.125	0.219	0.423	0.577
The OBJECT NOUN of SUBJECT	0.868	0.154	0.261	0.823	0.942

Table 10: RoBERTa model's performance on classes.

Class	Precision	Recall	F1 Score	PR AUC	ROC AUC
The NOUN by SUBJECT	0.977	0.999	0.988	0.981	0.435
SUBJECT's NOUN	0.878	0.937	0.906	0.933	0.719
SUBJECT's NOUN of OBJECT	0.872	0.985	0.925	0.970	0.843
The NOUN by SUBJECT of OBJECT	0.863	0.988	0.921	0.968	0.837
The NOUN of OBJECT	0.820	0.984	0.895	0.880	0.620
SUBJECT's OBJECT NOUN	0.686	0.947	0.795	0.839	0.693
The OBJECT NOUN by SUBJECT	0.663	0.943	0.778	0.828	0.700
The OBJECT NOUN	0.655	0.935	0.771	0.802	0.665
OBJECT's NOUN	0.660	0.796	0.722	0.805	0.704
The SUBJECT NOUN	0.617	0.674	0.644	0.651	0.519
The SUBJECT NOUN of OBJECT	0.497	0.569	0.530	0.605	0.524
The SUBJECT OBJECT NOUN	0.511	0.604	0.554	0.629	0.591
The NOUN of SUBJECT	0.832	0.373	0.515	0.764	0.824
OBJECT's NOUN of SUBJECT	0.936	0.349	0.508	0.553	0.680
The OBJECT NOUN of SUBJECT	0.933	0.686	0.790	0.831	0.930

Table 11: DeBERTa model's performance on classes.

Class	Precision	Recall	F1 Score	PR AUC	ROC AUC
The NOUN by SUBJECT	0.977	0.999	0.988	0.996	0.867
SUBJECT's NOUN	0.873	0.996	0.930	0.913	0.646
SUBJECT's NOUN of OBJECT	0.950	0.965	0.957	0.957	0.905
The NOUN by SUBJECT of OBJECT	0.936	0.963	0.949	0.983	0.912
The NOUN of OBJECT	0.882	0.978	0.927	0.940	0.802
SUBJECT's OBJECT NOUN	0.718	0.832	0.771	0.686	0.625
The OBJECT NOUN by SUBJECT	0.746	0.828	0.785	0.695	0.666
The OBJECT NOUN	0.690	0.825	0.752	0.650	0.582
OBJECT's NOUN	0.732	0.738	0.735	0.886	0.818
The SUBJECT NOUN	0.567	0.685	0.621	0.623	0.492
The SUBJECT NOUN of OBJECT	0.557	0.628	0.591	0.662	0.586
The SUBJECT OBJECT NOUN	0.539	0.526	0.532	0.571	0.528
The NOUN of SUBJECT	0.716	0.224	0.342	0.671	0.778
OBJECT's NOUN of SUBJECT	0.988	0.323	0.487	0.509	0.601
The OBJECT NOUN of SUBJECT	0.990	0.348	0.516	0.817	0.876

Table 12: Baseline and highest F_1 scores of classes.

Deverbal Nominal	Baseline F_1 Score	Highest F_1 Score
The NOUN by SUBJECT	**0.974**	**0.988**
SUBJECT's NOUN	0.936	0.930
SUBJECT's NOUN of OBJECT	**0.930**	**0.957**
The NOUN by SUBJECT of OBJECT	**0.925**	**0.949**
The NOUN of OBJECT	**0.918**	**0.927**
SUBJECT's OBJECT NOUN	0.907	0.795
The OBJECT NOUN by SUBJECT	**0.773**	**0.785**
The OBJECT NOUN	**0.742**	**0.752**
OBJECT's NOUN	0.742	0.735
The SUBJECT NOUN	0.701	0.652
The SUBJECT NOUN of OBJECT	0.645	0.591
The SUBJECT OBJECT NOUN	**0.422**	**0.569**
The NOUN of SUBJECT	**0.387**	**0.515**
OBJECT's NOUN of SUBJECT	**0.198**	**0.508**
The OBJECT NOUN of SUBJECT	**0.198**	**0.790**

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

Deverbal nominals are nominals that are derived from verbs and have noun phrase internal structures, however they also have argument taking ability like their verbal counterparts. Unique in many ways, they seem unpredictable morphologically, semantically and syntactically.

Chomsky shows that deverbal nominal structures cannot be linked to sentences by transformations. He hypothesizes that the determiner information of nominalization lies in the lexicon and can be expressed as fixed selectional and strict subcategorizational features of lexical elements. Grimshaw observes that only complex event nominals which have temporal organization shows argument taking ability like verbs. Although it is shown that the relationship between argument taking ability and event structure is not as straightforward as assumed, it is believed that there is still merit in following a lexicalist approach. To this end, this study presents a data-driven approach to predicting syntactic structures of deverbal nominals.

NOMLEX is a dictionary of deverbal nominal syntactic structures: it specifies syntactic structures of deverbal nominals produced, given the verb and the sentence structure the verb is used in. According to NOMLEX, 941 verbs used in 71 sentence structures can produce 3025 unique deverbal nominal syntactic structures and more than one deverbal nominal syntactic structure can be produced from a verb used in a sentence structure. In light of these, the problem is regarded as a multi-label classification problem and the information in NOMLEX is used to create the classes.

Deep learning models have the advantage of alleviating the feature engineering problem that is present in traditional machine learning techniques, and pre-trained deep learning models have the advantage of learning common language models from large amounts of data compared to training only with the task specific dataset. Transformer encoder architectures perform best in NLP classification tasks. On this account, transformer encoder based pre-trained deep learning models with different architectures and pre-training are used in the study.

Pre-trained deep learning models have their limitations: they cannot accept features and it is assumed that they cannot learn in depth syntactic knowledge from unsupervised learning. As a result, syntactic structures of sentences deverbal nominals are produced from and the semantic relation to their verbal counterparts are controlled in the study. It is observed that even with these restrictions, 702 verbs present in NOMLEX can produce 138 unique deverbal syntactic structures.

In order to leverage the common language representations pre-trained models learn from large amounts of corpora, available sentences in Gigaword and Wikipedia are used in the learning. Owing to this, information of selectional preferences of verbs are provided to the model in the fine-tuning stage implicitly. For parsing and filtering the data sources, EasyCCG parser is used.

Stratification is done on the verb level, meaning sentence examples of a particular verb is found only in the training set or the test dataset. Verbs with less than 1000 available sentences are excluded from the dataset and only 5000 of the sentences are included in the dataset for verbs that have more than 5000 sentences. Only the most produced classes, which are the classes that include only the prepositional phrases headed by "of" and "by" are included in the dataset. The resulting dataset consisted of 131 verbs and 362.387 sentences. These sentences can map into 15 classes that are not mutually exclusive.

For learning, four state-of-the-art transformer based pre-trained deep learning models with varying architectures and pre-training tasks are fine-tuned with the same dataset: BERT, XLNet, RoBERTa and DeBERTa. The models are evaluated with both sample-based and label-based evaluation metrics.

It is observed that in sample based evaluation techniques, RoBERTa and DeBERTa models have the same F_1 score, however the RoBERTa models is more sensitive and less specific compared to the DeBERTa model. DeBERTa model outperforms RoBERTa model in accuracy and hamming loss.

Label-based performance metrics are lower than sample-based performance metrics where they are comparable, which can be explained by the fact that since label-based performance metrics are averaged over labels, they give more weight to labels that are less represented in the dataset compared to sample-based performance metrics. It is a possibility that models fail to generalize or overgeneralize on less prevalent labels because of lack of representation.

When models' performance is evaluated on classes, it is seen that models' learning capabilities varies greatly over classes. It is observed that at least one of the models outperform baseline F_1 score in 10 of the 15 classes present. Among the unsuccessfully learned classes are the syntactic structures where only the subject or the object is realized in the possessive determiner position, two of the three classes that realize the subject in the prenoun noun modifier position and two of the three classes that realize the subject in the possessive determiner position. These observations suggest that determiners of an argument's realization in a syntactic position is dependent on both the argument and the syntactic position in question along with realization of other arguments in the deverbal nominal. In light of these, it is concluded that determiners of an argument's realization in a syntactic position is dependent on both the argument and the syntactic position in question along with realization of other arguments. The presented method was able to learn these determiners to various extents.

Because sentence examples of a particular verb is included in the validation set or the training set, the validation set consisted of sentence examples of verbs that are never seen by the model in the learning phase. By design, the study assumes that verbs are organized among themselves by deverbal syntactic structures they produce and any findings of the study would also confirm this organization.

7.2 Future Work

This study focuses on deverbal nominal syntactic structures that are derived from only one sentence structure. According to NOMLEX, verbs can produce deverbal nominals when present in 71 sentence structures and the study can be extended to include deverbal nominal syntactic structures that are derived from these sentence structures.

Only a subset of verbs that are known to produce deverbal nominals are included in the study because adequate number of sentence examples for them are not found in the studied data sources and more data sources can be included in the study.

In the study, four pre-trained deep learning models are fine-tuned. Unsupervised learning has been used to train these models and the study can be repeated with models that incorporate syntactic and semantic information.

Bibliography

Alexiadou, A. (2011). Statives and nominalization. *Recherches linguistiques de Vincennes*(40), 25–52.

Anderson, M. (1978). Np pre-posing in noun phrases. In *North east linguistics society* (Vol. 8, p. 3).

Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. *The nature of explanation in linguistic theory*, *31*, 67.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184–221). Waltham, Mass.: Ginn.

Chomsky, N. (2014). *Aspects of the theory of syntax* (Vol. 11). MIT press.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, *abs/1810.04805*. Retrieved from `http://arxiv.org/abs/1810.04805`

Diao, S. (2021). mlconjug3. *GitHub. Note: https://github.com/SekouDiaoNlp/mlconjug3 Cited by*.

Doron, E., & Rappaport-Hovav, M. (1991). Affectedness and externalization. In *North east linguistics society* (Vol. 21, p. 7).

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation.

Gildea, D., & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational linguistics*, *28*(3), 245–288.

Graff, D., Kong, J., Chen, K., & Maeda, K. (2003). English gigaword. *Linguistic Data Consortium, Philadelphia*, *4*(1), 34.

Grimshaw, J. (1990). *Argument structure.* the MIT Press.

Grover, C., Lascarides, A., & Lapata, M. (2005). A comparison of parsing technologies for the biomedical domain. *Natural Language Engineering*, *11*(1), 27–65.

Gurevich, O., & Waterman, S. A. (2009). Mapping verbal argument preferences to deverbals. In *2009 ieee international conference on semantic computing* (pp. 17–24).

He, P., Liu, X., Gao, J., & Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.* (To appear)

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *language*, *39*(2), 170–210.

Lapata, M. (2002). The disambiguation of nominalizations. *Computational Linguistics*, *28*(3), 357–388. Retrieved from https://aclanthology.org/J02-3004 doi: 10.1162/089120102760276018

Lees, R. B. (1962). The grammar of english nominalizations. *Journal of Symbolic Logic*, *27*(2).

Lewis, M., & Steedman, M. (2014). A* ccg parsing with a supertag-factored model. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 990–1000).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Macleod, C., Grishman, R., Meyers, A., Barrett, L., & Reeves, R. (1998). Nomlex: A lexicon of nominalizations. In *Proceedings of euralex* (Vol. 98, pp. 187–193).

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).

McCarthy, D., & Carroll, J. (2003). Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, *29*(4), 639–654.

Metheniti, E., Van de Cruys, T., & Hathout, N. (2020). How relevant are selectional preferences for transformer-based language models? In *Proceedings of the 28th international conference on computational linguistics* (pp. 1266–1278).

Newmeyer, F. (2009). Current challenges to the lexicalist hypothesis. *Time and again: Theoretical perspectives on formal linguistics*, 91–117.

Nicholson, J., & Baldwin, T. (2005). Statistical interpretation of compound nominalisations. In *Proceedings of the australasian language technology workshop 2005* (pp. 152–159).

Ortman, M. (2018). *Wikipedia sentences.* (data retrieved from https://www.kaggle.com/datasets/mikeortman/wikipedia-sentences)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, *63*(10), 1872–1897.

Reeves, R., Macleo, C., & Meyers, A. (1999). Manual of nomlex: n the regularized versio.

Resnik, P. S. (1993). Selection and information: A class-based approach to lexical relationships. *IRCS Technical Reports Series*, 200.

Roberts, W., & Egg, M. (2014). A comparison of selectional preference models for automatic verb classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 511–522).

Roeper, T. (1993). Explicit syntax in the lexicon: The representation of nominalizations. In *Semantics and the lexicon* (pp. 185–220). Springer.

Rozwadowska, B. (2017). Derived nominals. *The Wiley Blackwell Companion to Syntax, Second Edition*, 1–43.

Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Retrieved from `http://dx.doi.org/10.18653/v1/D15-1044` doi: 10.18653/v1/d15-1044

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), e0118432.

Schapire, R. E., & Singer, Y. (2000). Boostexter: A boosting-based system for text categorization. *Machine learning*, *39*(2), 135–168.

Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). On the stratification of multi-label data. *Machine Learning and Knowledge Discovery in Databases*, 145–158.

Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, *18*, 1–25.

Steedman, M., & Baldridge, J. (2011). Combinatory categorial grammar. *Non-Transformational Syntax: Formal and Explicit Models of Grammar. Wiley-Blackwell*, 181–224.

Sun, B., Wang, B., Che, W., Wu, D., Chen, Z., & Liu, T. (2022). *Improving pre-trained language models with syntactic dependency prediction task for chinese semantic error recognition.* arXiv. Retrieved from `https://arxiv.org/abs/2204.07464` doi: 10.48550/ARXIV.2204.07464

Szymański, P., & Kajdanowicz, T. (2017, February). A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*.

Szymański, P., & Kajdanowicz, T. (2017). A network perspective on stratification of multi-label data. In L. Torgo, B. Krawczyk, P. Branco, & N. Moniz (Eds.), *Proceedings of the first international workshop on learning with imbalanced domains: Theory and applications* (Vol. 74, pp. 22–35). ECML-PKDD, Skopje, Macedonia: PMLR.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, *3*(3), 1–13.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., … Bowman, S. (2019). Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*, *32*.

Wilks, Y. (1975). An intelligent analyzer and understander of english. *Communications of the ACM*, *18*(5), 264–274.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … Rush, A. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations* (pp. 38–45). Online: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2020.emnlp-demos.6` doi: 10.18653/v1/2020.emnlp-demos.6

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, *32*.

APPENDIX A

EXAMPLES OF SENTENCE STRUCTURES IN NOMLEX

A.1 NOM-NP

- The boy prefers cupcakes.
- He modified the plan.

A.2 NOM-NP-PP

- They attributed the painting to Masaccio.
- He acquired the painting for $2000

A.3 NOM-INTRANS

- He disappeared.
- He answered.

A.4 NOM-PP

- He advised against the compromise.
- She acted upon orders.

A.5 NOM-NP-AS-NP

- They accepted him as a doctor.
- He advertises her as the best hairdresser in the state.

A.6 NOM-PP-PP

- Nabisco competed against Nestles for market share.
- Nabisco competed for market share against Nestles.

A.7 NOM-THAT-S

- He observed that the world is better today.
- He argued that aliens do not exist.

A.8 NOM-S

- She knows that John is an "A" student.
- She knows John is an "A" student.

A.9 NOM-INTRANS-RECIP

- The streets intersect.
- They argued.

A.10 NOM-WH-S

- I wonder whether he is referring to Jake.
- I wonder if he is sick.

A.11 NOM-NP-TO-INF-OC

- We designated Allie to drive us home from the party.
- They contracted Smith & Co to remodel the basement.

A.12 NOM-POSSING

- I suggested my/his taking out loans to pay for college.
- I suggested taking out loans to pay for college.

A.13 NOM-HOW-S

- They complicated how it was done.
- He analyzed how the watch was working.

A.14 NOM-TO-INF-SC

- I want to go.
- He needs to win every argument.

A.15 NOM-P-POSSING

- Jay argued against their paying for the damage.
- Jay argued against paying for the damage.

A.16 NOM-NP-TO-NP

- The department allocates new students a computer
- The department allocates a computer to new students.
- The department allocates to new students the most expensive state-of-the art computers on the market.

A.17 NOM-NP-P-ING-OC

- I cautioned him about going.
- The state imprisoned the congressman for failing to pay taxes.

A.18 NOM-AS-NP

- Lulu failed as a pastry cook.
- He is known as the leader of the group.

A.19 NOM-NP-P-POSSING

- We collected money for Vi’s sweeping the road.
- We collected money for sweeping the road.

A.20 NOM-P-ING-SC

- They failed in attempting the climb.
- They confessed to cheating on the exam

A.21 NOM-PP-THAT-S

- They admitted to the authorities that they had smoked.
- The president agreed with his advisors that now would not be a good time to talk to the press.

A.22 NOM-P-WH-S

- He inquired about what they wanted to do.
- He inquired about what to do.

A.23 NOM-NP-ADVP

- He put it there.
- They treated them well.

A.24 NOM-ING-SC

- The police department continued accepting bribes.
- They delayed their/*his swimming the channel.

A.25 NOM-NP-PP-PP

- They converted the interest from US dollars to Swiss francs.
- They converted the interest to Swiss francs from US dollars.

A.26 NOM-ADVP

- He behaved badly.
- They settled there.

A.27 NOM-S-SUBJUNCT

- I demanded that he be in tune.
- I suggest (that) John come early.

A.28 NOM-FOR-TO-INF

- Sue wish for him to leave early
- I campaigned for her to become president.

A.29 NOM-PP-HOW-TO-INF

- She demonstrated to him how to do it.
- She demonstrated how to do it.
- They demonstrated to me how they sailed.
- They demonstrated how they sailed.

A.30 NOM-PART-PP

- I called out for help.
- I came back for her.

A.31 NOM-NP-ING-OC

- I caught John stealing.
- I imitated the president denying the charges.

A.32 NOM-EXTRAP-NP-S

- It annoyed him that no one answered.
- That no one answered annoyed him.

A.33 NOM-NP-AS-ADJP

- They characterized the play as well-acted.
- The society considers her as successful

A.34 NOM-P-NP-ING

- They worried over him drinking so much.
- They report the manager leaving the site.

A.35 NOM-NP-S

- I reminded her that the car had been stolen.
- I reminded her the car had been stolen.

A.36 NOM-PP-WH-S

- They mentioned to me what needed to be done.
- They concealed from him whether they would attack.
- He asked of everybody if he could have a piece or not.

A.37 NOM-PP-P-WH-S

- I argued with him about whether he should kill them.
- I argued with them about what to do.

A.38 NOM-NP-FOR-NP

- The chef prepared breakfast for the guest.
- The chef prepared the guest breakfast.
- The chef prepared for the guest a feast so magnificent that it became a national legend.

A.39 NOM-NP-ING

- I kept them laughing.
- John justified Max cheating.

A.40 NOM-NP-AS-ING

- She diagnosed him as being ill with the measles.
- HE identified him as being one of the victims.

A.41 NOM-NP-AS-NP-SC

- They served the king as messengers.
- He judged the facts as a scientist.

A.42 NOM-NP-P-WH-S

- I asked him about whether we would meet.
- John briefed me on whether to take the northern route.

A.43 NOM-PART

- I took over.
- He came back.

A.44 NOM-NP-P-ING-SC

- I spent time on classifying words.
- I modified my car by changing by wheel rims.

A.45 NOM-PART-NP-PP

- I counted down the minutes from ten.
- I restored back the room to its old version

A.46 NOM-PART-NP

- I fired off thim.
- I lifted off the dirt.

A.47 NOM-NP-NUNITP-TO-RANGE

- They reduced the price to $102 per share from $100.
- They reduced the price to $102 from $100 per share.

A.48 NOM-NP-NP

- The accident cost me $8000.
- She increased the wages 30 percent.

A.49 NOM-NP-AT-NUNITP-PRED

- They projected its value at 10 dollars a share.
- They put the price at $300 an ounce.

A.50 NOM-NUNITP-TO-RANGE

- The stock rose from $100 to $102
- The price rose to $102 from $100.

A.51 NOM-NP-ADJP-PRED

- He needs him healty.
- I imagine him happy.

A.52 NOM-PP-P-POSSING

- Jake argued with Mick about Clinton's visiting China.
- They argued among themselves about Clinton's visiting China.

A.53 NOM-P-NP-TO-INF-VC

- She appealed to him to leave the compound.
- He contracted with them to win the contest.

A.54 NOM-NP-TOBE

- Meteorologists predicted this year's winter to be the coldest of the decade.

A.55 NOM-NP-P-NP-ING

- I asked him about no one having been there.
- I asked him about there having been no witnesses.

A.56 NOM-POSSING-PP

- She attributed his giving up smoking to will power.
- She attributed giving up smoking to will power.
- She attributed to will power his giving up smoking.
- She attributed to will power giving up smoking.

A.57 NOM-NP-P-ING

- I prevented the child from running outside.
- I prohibited him from smoking.

A.58 NOM-ADVP-PP

- He resides here in Oxford.
- It boded ill for him.

A.59 NOM-P-NP-TO-INF-OC

- He motioned to John to eat the spinach.

A.60 NOM-NP-PP-AS-NP

- They mentioned the phone-call to me as a possible lead.
- She recommended him to me as a chiropractor.

A.61 NOM-PP-FOR-TO-INF

- They arranged with her for Johnny to take the bus to school.

A.62 NOM-ADJP-PRED-RS

- He looks good.

A.63 NOM-NP-TO-INF-SC

- John promised Mary to repair the desk lamp.

A.64 NOM-NUNITP-FROM-RANGE

- The stock price varied from twenty dollars to forty dollars.
- The prices for sofas range from $400 to $200.

A.65 NOM-P-NP-TO-INF

- He relies on Joan to come.

A.66 NOM-NP-WH-S

- He asked me whether the world is round.
- They asked him what to do.

A.67 NOM-WORDS

- We should answer Yes.
- The dog obeys Stop Now.
- He understands "No".

A.68 NOM-PP-THAT-S-SUBJUNCT

- They suggested to him that he be on time.

A.69 NOM-PART-AS-NP

- I took over as CEO.

A.70 NOM-PART-NP-AS-NP

- I took over the company as CEO.

A.71 NOM-NP-NUNITP-FROM-RANGE

- They extended the property from twenty feet to forty feet.

APPENDIX B

LEARNING GRAPHS

In this section, Precision vs. Recall and ROC curves of two best performing models are presented. Encodings for derived nominals used in graphs are presented in Table 13.

Table 13: Derived nominal encodings used in graphs.

Derived Nominal Syntactic Structure	Encoding
SUBJECT’s NOUN of OBJECT	0
The NOUN by SUBJECT of OBJECT	1
OBJECT’s NOUN	2
The NOUN of OBJECT	3
SUBJECT’s NOUN	4
The NOUN by SUBJECT	5
The SUBJECT OBJECT NOUN	6
The SUBJECT NOUN of OBJECT	7
SUBJECT’s OBJECT NOUN	8
The OBJECT NOUN by SUBJECT	9
The OBJECT NOUN	10
The SUBJECT NOUN	11
OBJECT’s NOUN of SUBJECT	12
The OBJECT NOUN of SUBJECT	13
The NOUN of SUBJECT	14

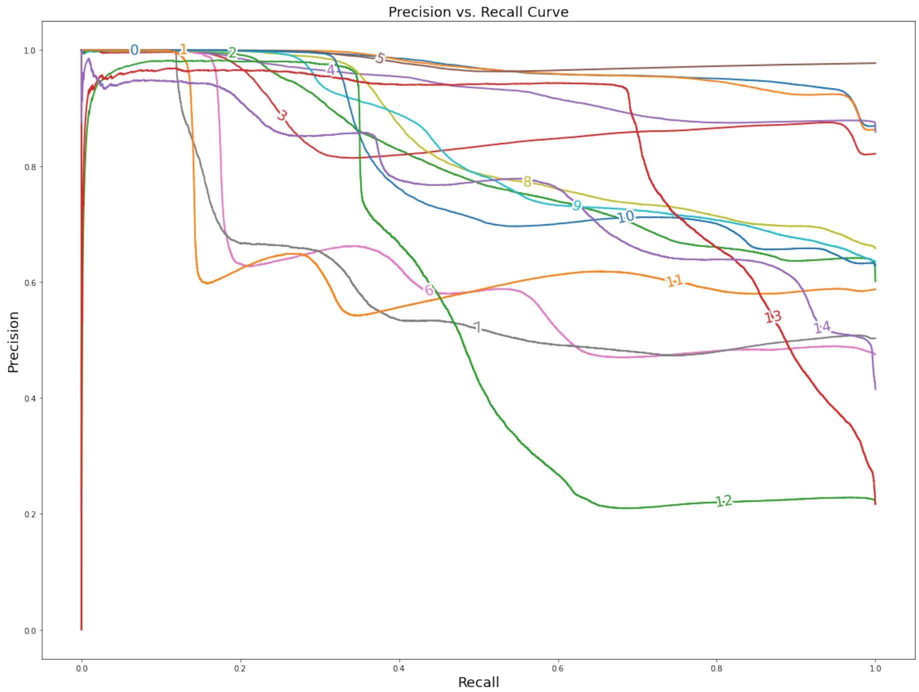

Figure 1: Precision vs. Recall Curve for BERT model.

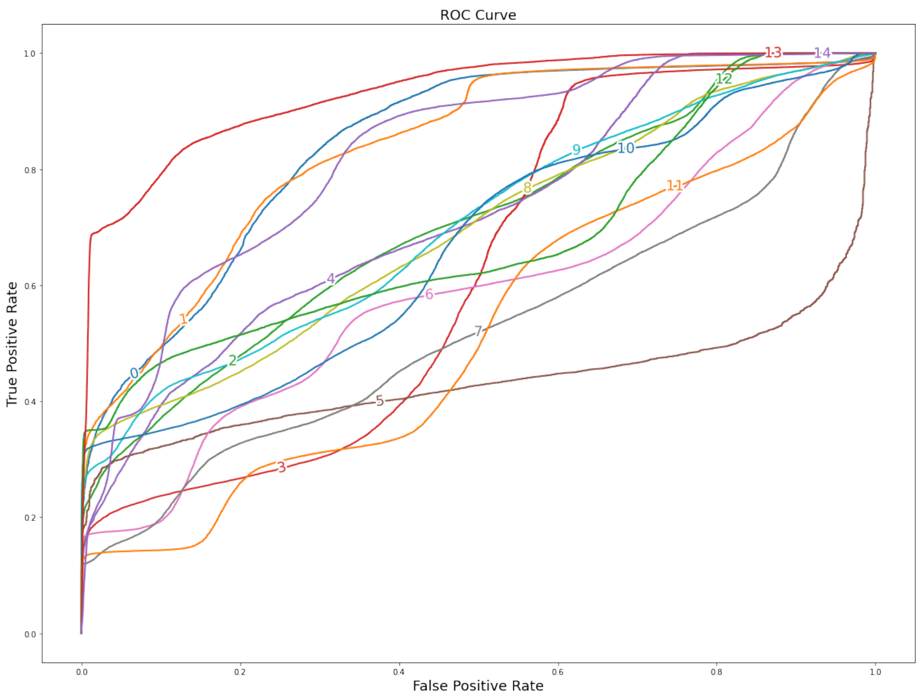

Figure 2: ROC Curve for BERT model.

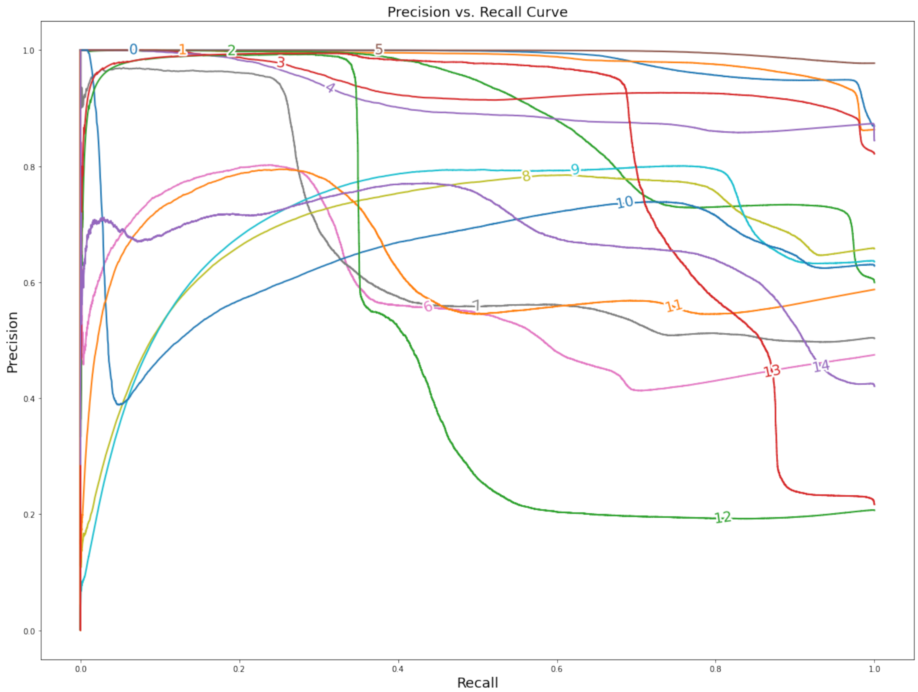

Figure 3: Precision vs. Recall Curve for XLNet model.

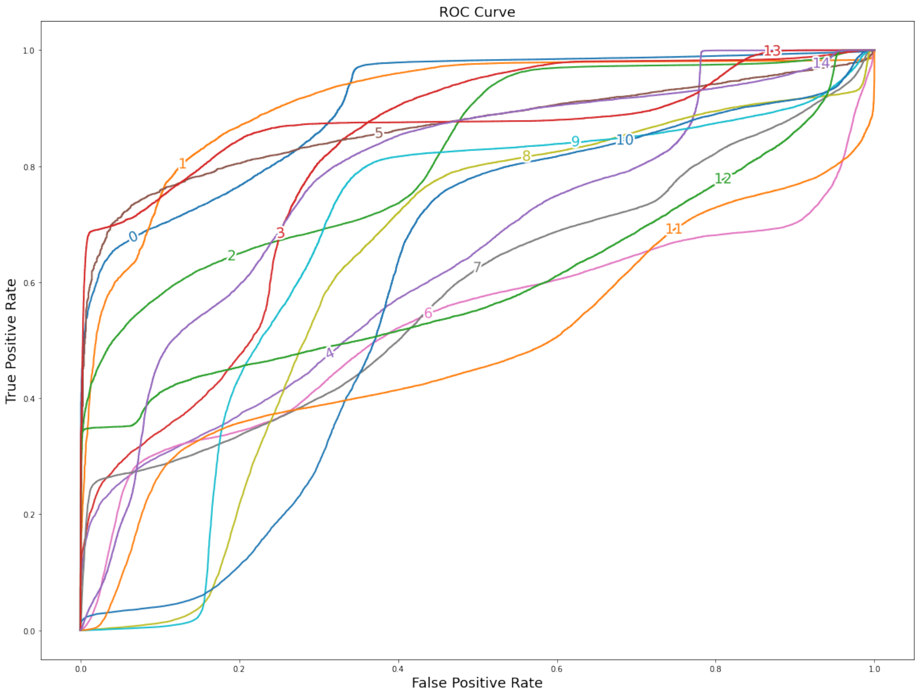

Figure 4: ROC Curve for XLNet model.

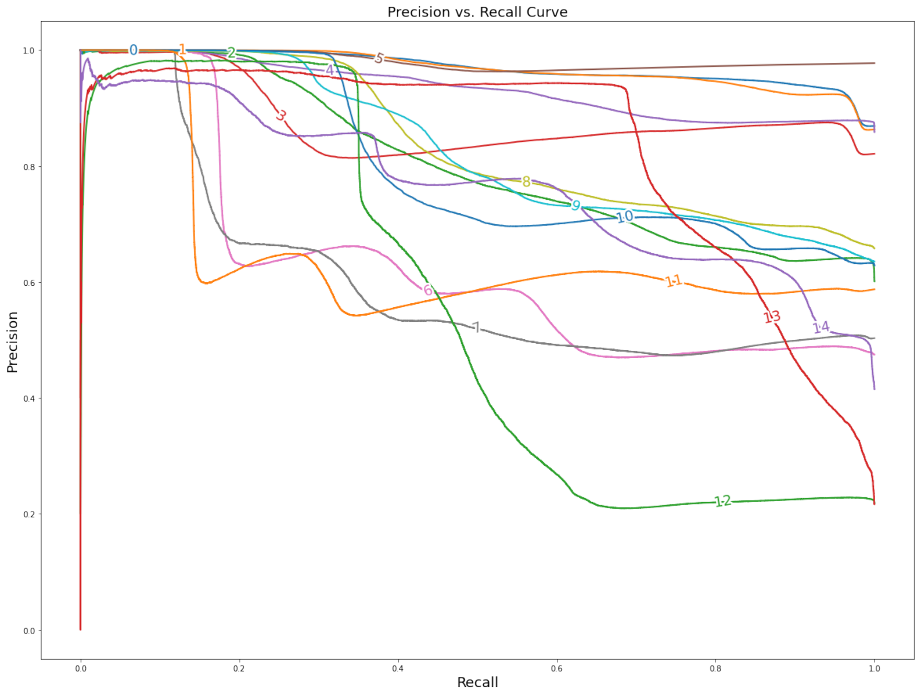

Figure 5: Precision vs. Recall Curve for RoBERTa model.

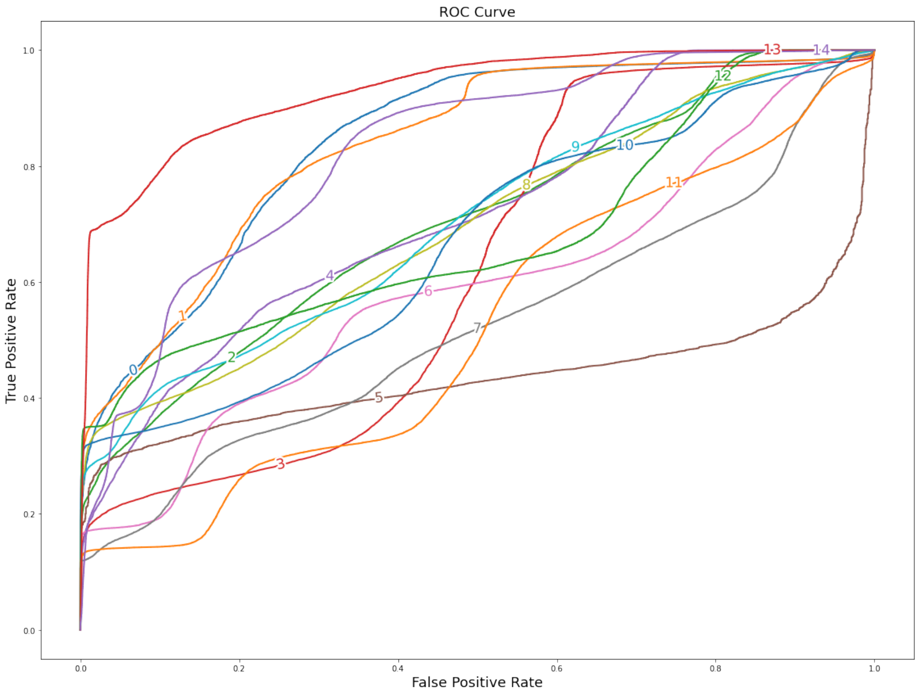

Figure 6: ROC Curve for RoBERTa model.

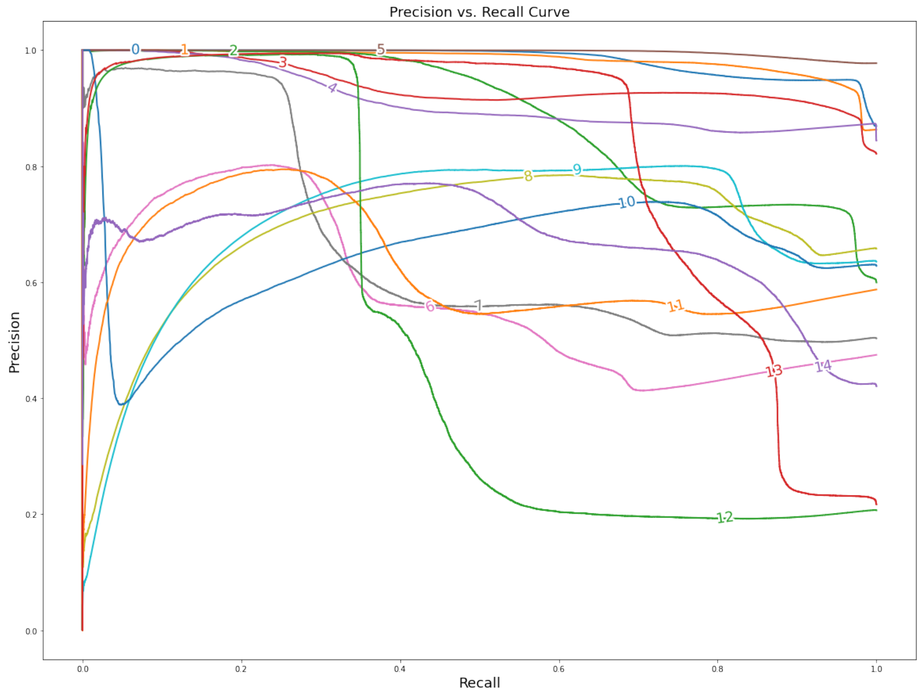

Figure 7: Precision vs. Recall Curve for DeBERTa model.

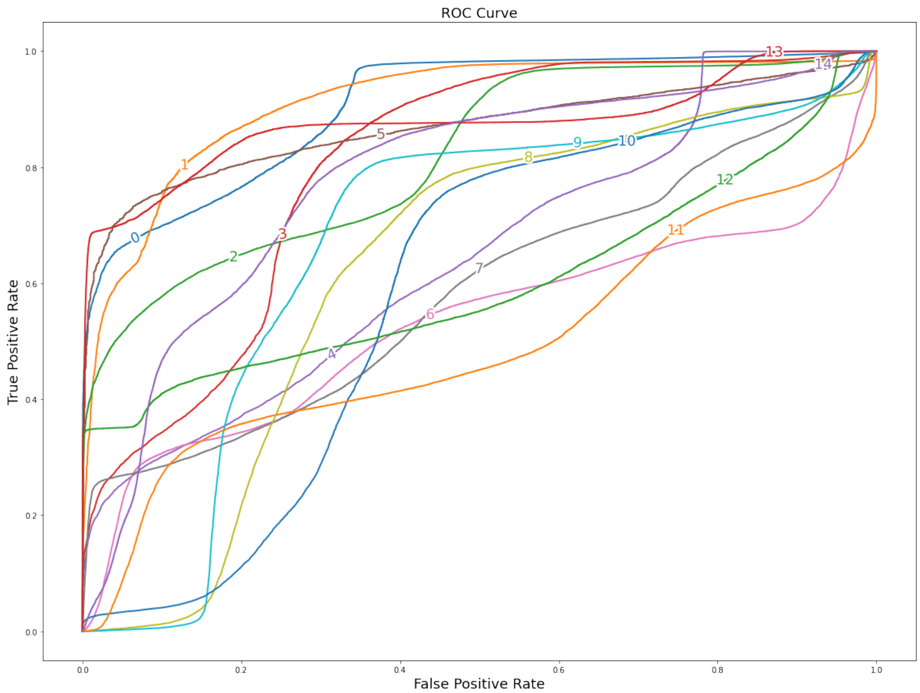

Figure 8: ROC Curve for DeBERTa model.