

***ENTAILMENT INFERENCE* UNTUK BAHASA INDONESIA  
DAN BAHASA JAWA NGOKO**

**PROPOSAL TESIS**

**Karya tulis sebagai salah satu syarat  
kelulusan MK IF5099 Metodologi Penelitian/Tesis 1**

**Oleh  
JALALUDDIN AL MURSYIDY FADHLURRAHMAN  
NIM: 23521059  
(Program Studi Magister Informatika)**



**INSTITUT TEKNOLOGI BANDUNG  
Mei 2023**

***ENTAILMENT INFERENCE* UNTUK BAHASA INDONESIA  
DAN BAHASA JAWA NGOKO**

Oleh  
**Jalaluddin Al Mursyidy Fadhlurrahman**  
**NIM: 23521059**  
**(Program Studi Magister Informatika)**

Institut Teknologi Bandung

Menyetujui  
Calon Tim Pembimbing


Tanggal 19 Mei 2023

Calon Pembimbing I

---

(Dr. Eng. Ayu Purwarianti, S.T., M.T.)

Calon Pembimbing II



---

(Alham Fikri Aji, Ph.D.)

## DAFTAR ISI

HALAMAN PENGESAHAN	i
DAFTAR ISI	ii
DAFTAR GAMBAR DAN ILUSTRASI	iii
DAFTAR TABEL	iv
DAFTAR SINGKATAN DAN LAMBANG	v
Bab I Pendahuluan	1
I.1 Latar Belakang	1
I.2 Masalah Penelitian	2
I.3 Tujuan Penelitian	2
I.4 Batasan Masalah	3
Bab II Tinjauan Pustaka	4
II.1 Natural Language Inference	4
II.1.1 Recognizing Textual Entailment	5
II.2 Teknik-teknik NLI	5
II.2.1 Word Embedding	5
II.2.2 Sentence Embedding	7
II.2.3 Model Kontekstual	8
II.2.4 <i>Transfer Learning</i>	9
II.3 <i>Low Resource Cross-Lingual NLI</i>	11
Bab III Metodologi	13
III.1 Linimasa Pengerjaan Tesis	13
III.2 <i>Dataset</i> NLI Bahasa Indonesia-Bahasa Jawa	13
III.2.1 Persyaratan Peserta Anotator	14
III.2.2 Metode Evaluasi <i>Dataset</i>	14
III.2.3 Hubungan Semantik Antar Kalimat	15
III.2.4 Pemilihan Label Emas	15
III.2.5 Perbaikan Teks Data	16
III.3 Metode yang Diajukan	16

## DAFTAR GAMBAR DAN ILUSTRASI

Gambar II.1	Skema metode <i>transfer learning</i> untuk melatih arsitektur <i>baseline</i> (Bandyopadhyay et al., 2022).	10
Gambar III.1	Skema metode transfer learning yang diajukan.	17

## DAFTAR TABEL

Tabel II.1 Contoh pasangan kalimat premis-hipotesis beserta label.	4
Tabel III.1 Linimasa rencana pengerjaan tesis.	13

## DAFTAR SINGKATAN DAN LAMBANG

SINGKATAN	Nama	Pemakaian pertama kali pada halaman
NLI	<i>Natural Language Inference</i>	1
NLU	<i>Natural Language Understanding</i>	1
MultiNLI	<i>Multi-Genre Natural Language Inference</i>	1
SNLI	<i>Stanford Natural Language Inference</i>	1
WReTE	<i>Wikipedia Recognizing Textual Entailment</i>	1
INARTE	<i>Indonesian Recognizing Textual Entailment</i>	1
RTE	<i>Recognizing Textual Entailment</i>	5
CCA	<i>Canonical Correlation Analysis</i>	6
CBOW	<i>Continuous Bag-of-Words</i>	7
MT	<i>Machine Translation</i>	7
BERT	<i>Bidirectional Encoder Representations from Transformers</i>	8
XLMR	<i>Cross Lingual Model-Roberta</i>	8
LAMBANG	Nama	Pemakaian pertama kali pada halaman
MS	<i>Meta training set</i>	9
$(S_i^s, T_i^s)$	<i>Set support dan set query dari masalah pembelajaran ke-i</i>	9
$N^s$	Jumlah masalah pembelajaran yang digunakan untuk <i>training</i>	9
MT	<i>Meta testing set</i>	9
$(S_i^t, T_i^t)$	<i>Set support dan set query dari masalah pembelajaran tes ke-i</i>	9
$N^t$	Jumlah masalah pembelajaran yang digunakan untuk <i>testing</i>	9
$F$	Matriks bobot pada <i>layer feed-forward</i>	10
$C$	Vektor konteks dari token CLS	10
$W$	Matriks bobot	10
$b$	Bias	10
$P$	Probabilitas label	11

## **Bab I   Pendahuluan**

Bab ini mencakup landasan, tujuan dan batasan masalah penelitian. Bab ini bertujuan untuk memberikan gambaran secara menyeluruh tentang penelitian, serta urgensi dan tujuan yang diharapkan dapat diraih dengan penelitian ini.

### **I.1    Latar Belakang**

*Natural Language Processing (NLP)* dalam beberapa tahun terakhir telah mengalami kemajuan yang pesat. Meskipun demikian, penelitian yang dilakukan berpusat pada bahasa Inggris (Hu et al., 2020). Menurut Hedderich et al. (2021), riset-riset yang telah dilakukan terhadap bahasa dengan sumber daya terbatas (*low resource language*), seperti bahasa Jawa, masih sangat minim, dan belum diketahui apakah model NLP dapat memahami bahasa Jawa. Oleh karena itu, *Natural Language Inference (NLI)* dieksplorasi karena NLI umum dipakai sebagai tolok ukur pemahaman model NLP terhadap konteks (Parikh et al., 2016), peringkasan teks (Falke et al., 2019; Kryści et al., 2020), pemahaman bacaan (Trivedi et al., 2019), dan sistem dialog (Welleck et al., 2019).

Keterbatasan data latih yang tersedia, kurangnya sumber daya linguistik seperti korpus teks, kamus, ataupun informasi struktur tata bahasa yang memadai (Upadhyay et al., 2016), serta penyesuaian dengan kebudayaan dan nuansa bahasa baru seperti adanya idiom, ungkapan-ungkapan, peribahasa, ataupun ekspresi emosi (Budur et al., 2020) dapat menjadi tantangan untuk mendukung pemrosesan bahasa alami ke dalam *domain* bahasa yang baru. Oleh karena itu, riset ini juga bertujuan untuk melihat kemampuan model untuk melakukan *transfer learning* dari bahasa yang serupa, yaitu bahasa Indonesia. Tidak adanya data yang tersedia untuk menguji NLI dalam bahasa Jawa, membuat pembangunan data NLI dalam bahasa Jawa perlu dilakukan.

## I.2 Masalah Penelitian

Berdasarkan yang sudah dijabarkan pada bagian latar belakang, permasalahan utama pada NLI lintas bahasa adalah teks-teks pada bahasa yang berbeda memiliki struktur semantik, sintaktis, serta tata bahasa yang berbeda (Bandyopadhyay et al., 2022). Aspek permasalahan lain pada penelitian NLI yang sudah ada adalah pasangan kalimat premis-hipotesis yang mengandung antonim, penalaran yang melibatkan angka, kata-kata yang saling tumpang tindih, adanya kata-kata negasi yang kuat, kalimat premis yang lebih panjang dari kalimat hipotesis, dan kesalahan pengejaan, *lexical overlap*, inferensi tanpa pengetahuan *lexical* dan dunia nyata, dan struktur sintaktis. Selain itu, pada NLI lintas bahasa, apabila bahasa yang digunakan merupakan bahasa yang *low resource*, maka representasi katanya tidak hanya kehilangan hubungan semantik, tapi juga hubungan *coarse-grained* dan induksi leksikonnya akan hilang (Zhang et al., 2016). Permasalahan selanjutnya pada NLI lintas bahasa, khususnya bahasa Indonesia-bahasa Jawa, model yang menjadi *baseline* belum mampu melakukan penalaran apabila terdapat ungkapan, peribahasa, atau ekspresi emosi dalam bahasa Jawa. Model masih cenderung mengkategorikan sebuah pasangan kalimat premis-hipotesis menjadi netral apabila terdapat ungkapan-ungkapan, peribahasa, atau ekspresi emosi dalam bahasa Jawa.

## I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan *entailment inference* terhadap pasangan kalimat premis-hipotesis berbahasa Indonesia-bahasa Jawa. Secara mendetail, penelitian ini memiliki tujuan untuk:

1. Membangun model yang mampu melakukan penalaran apabila terdapat ungkapan-ungkapan, peribahasa, atau ekspresi emosi dalam bahasa Jawa.
2. Membangun *dataset* NLI lintas bahasa untuk bahasa Indonesia dan bahasa Jawa. Pengembangan *dataset* dimaksudkan untuk menjawab tantangan tidak adanya data yang tersedia untuk NLI bahasa Jawa.



#### **I.4 Batasan Masalah**

Batasan masalah yang diterapkan pada penelitian ini adalah:

1. Membangun *task* NLI lintas bahasa antara bahasa Indonesia dan bahasa Jawa.
2. Bahasa Jawa yang digunakan adalah varian bahasa Jawa Timuran Ngoko.

## Bab II Tinjauan Pustaka

Bab tinjauan pustaka ini berisi uraian *state-of-the-art* penelitian mengenai NLI. Tinjauan pustaka dilakukan untuk mengkaji sejauh mana perkembangan penelitian untuk domain permasalahan NLI.

### II.1 Natural Language Inference

Penalaran dan pengambilan kesimpulan merupakan tugas yang penting dan mendasar pada *Natural Language Understanding* (Maccartney & Manning, 2008). *Natural Language Inference* (NLI) secara spesifik bertujuan untuk menentukan sebuah kalimat hipotesis  $h$  dapat disimpulkan dari kalimat premis  $p$ . Conneau et al. (2018) memberikan contoh pasangan kalimat hipotesis dan kalimat premis beserta labelnya sebagaimana yang dapat dilihat pada Tabel II.1. Contoh pasangan kalimat yang ditampilkan pada Tabel II.1 merupakan bentuk *recognizing textual entailment*.

Tabel II.1 Contoh pasangan kalimat premis-hipotesis beserta label.

Premis	Hipotesis	Label
<i>You can't stay here</i>	<i>You can leave</i>	<i>Entailment</i>
<i>Conceptually cream skimming has two basic dimensions - product and geography</i>	<i>Product and geography are what make cream skimming work</i>	<i>Neutral</i>
<i>At the end of Rue des Francs-Bourgeois is what many consider to be the city's most handsome residential square , the Place des Vosges , with its stone and red brick facades</i>	<i>Place des Vosges is constructed entirely of gray marble</i>	<i>Contradiction</i>

### II.1.1 Recognizing Textual Entailment

*Recognizing Textual Entailment* (RTE) merupakan tantangan pada NLI yang bertujuan untuk mengevaluasi kemampuan mesin untuk mendapatkan *background knowledge*, dan kemampuan penalaran yang dibutuhkan oleh manusia untuk menentukan apakah sebuah teks dapat disimpulkan dari teks lainnya (Dagan et al., 2006). Tantangan paling umum adalah sebuah sistem diberikan sebuah teks dan hipotesis. Sistem kemudian diharapkan mampu memprediksi apakah hipotesis dapat diambil dari teks tersebut. Pada tahun 2008 dan 2009, muncul tantangan lain yang merupakan *decision task* tiga arah dimana sistem diharapkan mampu mengenali hubungan kalimat kontradiksi antara sebuah teks dengan kalimat hipotesis (Bentivogli et al., 2009; Giampiccolo et al., 2008). Tantangan selanjutnya pada tahun 2010 dan 2011 berupa sebuah kalimat hipotesis yang dapat disimpulkan dari beberapa kalimat pada sebuah korpus (Bentivogli et al., 2010, 2011). Untuk menjawab tantangan NLI tersebut, dibutuhkan sumber pengetahuan sebagai basis data bagi model untuk dapat memberikan argumen.

## II.2 Teknik-teknik NLI

Untuk menyelesaikan task-task NLI, berbagai pendekatan telah dikembangkan. Bagian II.2 ini akan menjelaskan penelitian-penelitian yang telah dilakukan untuk melakukan inferensi bahasa alami.

### II.2.1 Word Embedding

Beberapa teknik telah diusulkan untuk melakukan NLI. Salah satu pendekatan yang diusulkan berada pada level kata, yaitu *multilingual word embedding*. Metode ini memerlukan data latih dalam bentuk leksikon dwibahasa kecil untuk menyelaraskan *embedding* sumber dan target ke dalam satu ruang *embedding* yang sama. Penelitian yang dilakukan oleh Mikolov et al. (2013) menerapkan teknik ini dengan cara mempelajari proyeksi linier antara ruang-ruang vektor yang merepresentasikan tiap-tiap bahasa. Metode ini terdiri dari dua langkah sederhana, yaitu membangun model monolingual dan menggunakan kamus dwibahasa kecil untuk mempelajari proyeksi linier antar bahasa. Pada saat pengujian, kata-kata pada korpus monolingual dapat diterjemahkan dengan memproyeksikan representasi

vektornya dari ruang bahasa sumber ke ruang bahasa target, kemudian vektor kata paling mirip dengan kata terjemahannya ditampilkan.

Pada penelitian lain yang dilakukan oleh Faruqui dan Dyer (2014), vektor hubungan linier antara sebuah kata dari bahasa sumber dengan bahasa target direpresentasikan menggunakan *Canonical Correlation Analysis* (CCA). CCA menghasilkan dua vektor proyeksi (satu untuk setiap variabel) yang memiliki korelasi yang paling optimal.

Metode-metode untuk mendapatkan *word embedding* lintas bahasa kebanyakan menggunakan korpus paralel yang besar (Gouws et al., 2015; Luong et al., 2015). Terdapat beberapa usulan yang meringankan kebutuhan akan data korpus yang besar, yaitu dengan menggunakan korpus yang menggunakan *document-aligned* atau *label-aligned* (Mogadala dan Rettinger, 2016; Søgaard et al., 2015). Akan tetapi, korpus ini tidak selalu tersedia untuk seluruh bahasa dan tetap memerlukan pengawasan dengan ribuan data dwibahasa. Penelitian yang dilakukan oleh Artetxe et al. (2017) menunjukkan bahwa *word embedding* antar bahasa dapat dihasilkan tanpa memerlukan pengawasan dengan data dwibahasa melebihi 25 pasang kata. Sementara, penelitian yang dilakukan oleh Zhang et al. (2016) menunjukkan bahwa *word embedding* bisa didapatkan dengan pengawasan terhadap data dwibahasa dengan menggunakan 10 pasang kata saja. Meskipun penelitian Zhang et al. (2016) berhasil menangkap hubungan *coarse-grained* antar kalimat, penelitian tersebut gagal pada induksi leksikon dwibahasa.

Metode *word embedding* memiliki beberapa kelemahan. Penggunaan korpus paralel yang besar menyebabkan bias gender, ras, stereotip budaya (Bolukbasi et al., 2016). Kelemahan selanjutnya, *word embedding* hanya dapat dibuat dari kata-kata yang sudah ada pada korpus latihnya saja (Mikolov, Chen, et al., 2013b). Apabila sebuah kata tidak ada pada korpus latih, maka kata tersebut tidak memiliki vektor *embedding*. Hal ini dapat menyebabkan model memiliki batasan domain. Kelemahan lainnya, *word embedding* susah untuk menafsirkan makna dari setiap *embedding*. Hal ini dapat menjadi masalah apabila model membuat kesalahan,

ataupun *embedding* digunakan pada domain yang mementingkan penafsiran (Li et al., 2018).

### II.2.2 Sentence Embedding

Pendekatan lain untuk melakukan NLI adalah dengan menggunakan pembelajaran dengan representasi kalimat (Arora et al., 2017; Blacoe dan Lapata, 2012; Wieting et al., 2016). Sebuah cara untuk mendapatkan *sentence embedding* adalah dengan melakukan rata-rata berbobot terhadap representasi kata-kata, yang sering disebut dengan *Continous Bag-of-Words* (CBOW) (Conneau et al., 2018). Cara lain untuk mendapatkan *sentence embedding* adalah dengan menggunakan model SkipThought, yang merupakan penelitian dari Kiros et al. (2015) dan kelanjutan dari model skip-gram hasil penelitian oleh Mikolov, Chen, et al. (2013a). Model SkipThought ini dapat menangkap dependensi semantik dan sintaktis pada representasi kalimat (Kiros et al., 2015). Penelitian oleh Bandyopadhyay et al. (2022) dan Howard dan Ruder (2018) menunjukkan bahwa penggunaan model bahasa pra-latih dapat bekerja dengan baik ketika *hidden state* digunakan sebagai vektor kontekstual, maupun ketika seluruh model digunakan untuk *fine-tuning*.

Pada penelitian yang dilakukan oleh Espana-Bonet et al. (2017) serta Schwenk dan Douze (2017), dilakukan representasi kalimat antar bahasa dengan melatih sistem *Machine Translation* (MT) *sequence-to-sequence* pada beberapa bahasa untuk mempelajari ruang *sentence embedding* lintas bahasa. Moritz Hermann dan Blunsom (2014) dalam penelitiannya mengusulkan sebuah model vektor untuk mempelajari representasi pada *document level* dengan menggunakan unigram dan bigram. Zhou et al. (2016) dalam penelitiannya mempelajari representasi dokumen dwibahasa dengan meminimalisasi jarak Euclid antara sebuah dokumen dengan terjemahannya.

Metode *sentence embedding* memiliki kelemahan, yaitu konteks yang terbatas. *Sentence embedding* dibentuk dari kata-kata dalam sebuah kalimat, tapi tidak menggambarkan konteks dari keseluruhan dokumen Reimers dan Gurevych, (2019). Selain itu, *sentence embedding* juga memiliki kesulitan memahami kalimat

yang panjang. Hal ini disebabkan oleh adanya informasi yang hilang ketika diaplikasikan pada kalimat yang panjang, terutama ketika informasi penting berada pada ujung awal atau akhir kalimat (Cer et al., 2018). Mirip dengan *word embedding*, menurut Conneau et al. (2017), *sentence embedding* juga biasanya dilatih pada korpus umum, yang tidak menangkap nuansa bahasa pada domain tertentu. Hal ini menyebabkan model tidak dapat bekerja dengan baik pada domain yang spesifik, seperti medis atau hukum. Cer et al. (2018) dalam penelitiannya menemukan bahwa kata-kata negasi sulit ditangkap dengan *sentence embedding* karena arti dari sebuah kalimat dapat berubah drastis tergantung sebuah kata dinegasikan atau tidak.

### II.2.3 Model Kontekstual

Penelitian-penelitian terkini pada NLI berfokus pada meningkatkan akurasi dari model dengan menggunakan informasi kontekstual. Model kontekstual menggunakan model bahasa pra-latih, seperti BERT, untuk menghasilkan representasi dari kata-kata dan kalimat yang memperhitungkan konteks di sekitarnya. Mahendra et al. (2021) menemukan dalam penelitiannya bahwa kinerja model terhadap tantangan NLI lintas bahasa, kinerja yang dihasilkan oleh BERT berada dibawah XLM-Roberta (XLMR). Hal ini terjadi dikarenakan ukuran data pra-latih pada XLMR memiliki porsi bahasa Indonesia yang lebih besar dibandingkan pada BERT.

Hu et al. (2020), Qi et al. (2022), Soudani et al. (2022) dan Tandon et al. (2021) dalam penelitiannya menggunakan XLMR sebagai model *baseline* memberikan hasil yang serupa dengan penelitian Mahendra et al. (2021), dimana XLMR memberikan hasil akurasi tertinggi. Hal ini dikarenakan XLMR melakukan *task* pembelajaran lintas-bahasa dengan menggunakan korpus paralel 42 GB. Akan tetapi, model kontekstual ini belum dapat memecahkan permasalahan klasik dalam domain NLI lintas-bahasa. Pada kasus *low-resource* NLI, permasalahan kurangnya pengetahuan model terhadap pengetahuan dunia nyata, kata negasi yang kuat, penalaran numerik, penalaran pada kalimat premis yang memiliki panjang lebih

dari 1 kalimat, kata-kata yang saling tumpang-tindih, serta penalaran secara umum, masih menjadi kelemahan teknik NLI dengan menggunakan model kontekstual.

#### II.2.4 *Transfer Learning*

Pendekatan lain yang dilakukan untuk NLI adalah dengan *transfer learning*. Penelitian terkini yang menggunakan metode *transfer learning* adalah penelitian oleh Soudani et al. (2022) yang menggunakan Meta-Learning. Meta-Learning merupakan salah satu metode *transfer learning* yang mempelajari model dengan mempelajari banyak permasalahan pembelajaran (Liu et al., 2020). Meta-Learning secara spesifik menggunakan *meta training set* seperti yang ditunjukkan oleh persamaan II.1.

$$MS = \{(S_i^s, T_i^s)\}_{i=1}^{N^s} \quad (II.1)$$

dengan:

$MS$  : *meta training set*

$(S_i^s, T_i^s)$  : *set support* dan *set query* dari masalah pembelajaran ke- $i$

$N^s$  : jumlah masalah pembelajaran yang digunakan untuk *training*

Meta-Learning juga menggunakan *meta test set* seperti yang ditunjukkan oleh persamaan II.2.

$$MT = \{(S_i^t, T_i^t)\}_{i=1}^{N^t} \quad (II.2)$$

dengan:

$MT$  : *meta testing set*

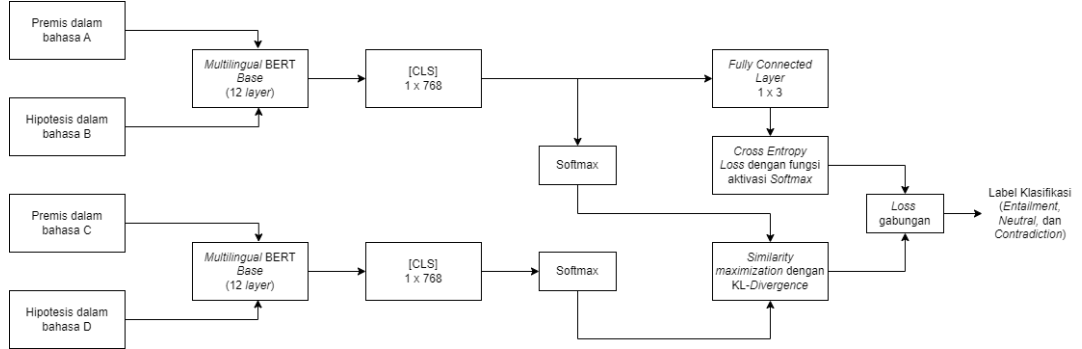
$(S_i^t, T_i^t)$  : *set support* dan *set query* dari masalah pembelajaran tes ke- $i$

$N^t$  : jumlah masalah pembelajaran yang digunakan untuk *testing*

Diketahui  $MS$ , *meta-learner* dapat mempelajari pemetaan pasangan  $(S, T)$  ke algoritma yang meningkatkan  $S$  untuk dapat secara optimal menyelesaikan  $T$ .

Dikarenakan kurangnya distribusi *task* dengan baik, meta-learning masih berupaya memperlakukan *dataset* sebagai *task*. Terlebih lagi, distribusi *task* secara ad-hoc menyebabkan masalah kualitas dan kuantitas.

Pendekatan *transfer learning* lain yang dilakukan oleh Bandyopadhyay et al. (2022) adalah dengan menggunakan Kullback-Leibler *Divergence* sebagai upaya untuk meningkatkan kemiripan distribusi probabilitas untuk pasangan pra-latih pada model guru dan distribusi probabilitas untuk pasangan latih pada model murid.



Gambar II.1 Skema metode *transfer learning* untuk melatih arsitektur *baseline* (Bandyopadhyay et al., 2022).

Pada skema metode *transfer learning* yang dilakukan oleh (Bandyopadhyay et al., 2022), seperti yang ditunjukkan oleh Gambar II.1, model guru merupakan model yang telah dilatih sebelumnya dalam pasangan bahasa C dan bahasa D, dan model murid merupakan model yang akan dilatih dengan pasangan bahasa A dan bahasa B. Masukan dari model berupa pasangan kalimat premis dan kalimat hipotesis. Setelah pasangan kalimat dimasukkan ke model, *hidden state* terakhir diambil untuk diekstrak token pertamanya. Token pertama dari *hidden state* terakhir ini merepresentasikan *token embedding* CLS pada BERT. Vektor konteks ini memiliki dimensi  $1 \times 768$ . Selanjutnya, representasi dari token CLS ini dimasukkan ke *layer feed-forward*. *Layer feed-forward* dapat direpresentasikan dengan persamaan II.3.

$$F = C \cdot W^T + b \quad (\text{II.3})$$

dengan:

- $F$  : bobot pada *layer feed-forward*
- $C$  : vektor konteks dari token CLS
- $W$  : matriks bobot
- $b$  : bias



Kemudian, probabilitas label dihitung dengan persamaan fungsi aktivasi Softmax yang ditunjukkan oleh persamaan II.4. Hasil perhitungan dari fungsi aktivasi ini memiliki dimensi sebesar  $1 \times 3$ . Semua parameter pada BERT dan *layer feed-forward* dilakukan *fine-tuning* secara bersama untuk memaksimalkan probabilitas logaritmik dari label yang benar dengan menggunakan fungsi *loss cross-entropy* standar.

$$P = \text{softmax}(F) \quad (\text{II.4})$$

dengan:

$P$  : probabilitas label

$F$  : bobot pada *layer feed-forward*.

### II.3 Low Resource Cross-Lingual NLI

Beberapa penelitian telah dilakukan untuk menggali teknik NLI lintas-bahasa yang minim sumber. Soudani et al. (2022) meneliti teknik NLI untuk bahasa Persia dengan menggunakan Meta-Learning. Pada penelitian ini, pasangan kalimat dalam 2 bahasa diberikan ke *meta-learner*. Algoritma Meta-Learning berupa *prototypical network* dan *reptile* digunakan pada keluaran dari model XLMR. Perbandingan kinerja antara dua algoritma Meta-Learning menghasilkan kinerja *prototypical network* yang lebih unggul. Hal ini dikarenakan *prototypical network* menggunakan representasi kelas, alih-alih representasi contoh. Meski penggunaan Meta-Learning meningkatkan kinerja model pada bahasa target, percobaan penambahan data target dengan data dari task tambahan untuk Meta-Learning menghasilkan akurasi yang menurun. Hal ini dikarenakan proporsi data target jauh lebih rendah daripada data tambahan. Penelitian ini juga mengemukakan permasalahan bahwa penambahan DReCa sebagai strategi augmentasi *task* dengan algoritma *prototypical network* tidak berpengaruh apabila *task* tambahan yang digunakan lebih dari 1 *task* tambahan. Selain itu, cara mendefinisikan strategi *task augmentation* yang tepat juga menjadi tantangan.

Hasil kinerja model NLI lintas-bahasa tertinggi secara umum ada pada bahasa daerah yang memiliki representasi paling besar pada korpora pra-latih aslinya. Kinerja model untuk bahasa daerah kemudian menurun seiring berkurangnya

representasi bahasa tersebut. Penelitian lain yang dilakukan oleh Flek et al. (2022), yakni NLI terhadap bahasa asli Amerika, menggunakan teknik *zero-shot* dan menggabungkannya dengan adaptasi model pra-latih, yaitu *continued pre-training* untuk mengatasi masalah sedikitnya representasi bahasa low resource pada model pra-latih. Penelitian didahului dengan menerjemahkan dataset XNLI bahasa Spanyol ke dalam 10 bahasa penduduk asli Amerika, yaitu Asháninka, Aymara, Bribri, Guarani, Nahuatl, Otomí, Quechua, Rarámuri, Shipibo-Konibo dan Wixarika. Model selanjutnya dilatih dengan adaptasi *continued pre-training*. Penggunaan *continued pre-training* memberikan kinerja yang paling andal untuk bahasa *low resource* karena tidak memperkenalkan parameter baru yang diinisialisasi secara acak ke model.

Pfeiffer et al. (2020) dalam penelitiannya membuktikan bahwa *continued pre-training* pada bahasa target secara monolingual dengan menggunakan *Masked Language Modeling* (MLM) dapat mengadaptasi *Multilingual Pretrained Language Model* (MPLM) secara efektif. Pendekatan lain oleh Jundi dan Lapesa (2022) mencoba untuk menggunakan sistem MT untuk menerjemahkan data berlabel pada bahasa sumber ke bahasa target, dan melakukan *fine-tuning* MPLM pada pasangan bahasa sumber dan bahasa target. Penelitian lain yang berusaha mengatasi permasalahan bahasa dengan sumber yang minim adalah dengan menggunakan leksikon beranotasi dwibahasa (Wang et al., 2022). Pada penelitian ini, leksikon antara bahasa Inggris dengan bahasa target diekstrak. Selanjutnya, data *pseudo* berbahasa tunggal untuk bahasa target dibuat dengan menggantikan kata-kata dalam bahasa Inggris dengan terjemahannya apabila kata tersebut ada pada leksikon dwibahasa. Apabila terdapat beberapa hasil terjemahan yang mungkin, sampel kata dalam bahasa target dipilih secara acak. Kemudian, model mBERT dilatih menggunakan data *pseudo* berbahasa tunggal dalam bahasa target. Selanjutnya, model mBERT tadi dilakukan *fine-tuning* untuk *downstream task* dengan menggabungkan data berlabel bahasa Inggris dan data *pseudo* berlabel.

## Bab III Metodologi

Bab ini akan membahas rencana metode, rancangan, dan perencanaan waktu untuk mengerjakan penelitian. Pembahasan meliputi pembuatan *dataset* NLI untuk bahasa Indonesia dan bahasa Jawa, serta metode yang diajukan.

### III.1 Linimasa Pengerjaan Tesis

Pelaksanaan tesis dilakukan dengan target untuk dapat selesai pada akhir semester Ganjil 2023/2024. Linimasa rencana pengerjaan tesis ditunjukkan oleh Tabel III.1.

Tabel III.1 Linimasa rencana pengerjaan tesis.

Kegiatan	Bulan di Tahun 2023											
	1	2	3	4	5	6	7	8	9	10	11	12
Studi literatur												
Analisis permasalahan												
Pembuatan proposal												
Pembuatan <i>dataset</i> uji												
Pengujian <i>baseline</i>												
Publikasi <i>paper</i> atau <i>conference</i>												
Pembuatan metode yang diajukan												
Pengujian metode yang diajukan												
Pembuatan laporan akhir												

### III.2 *Dataset* NLI Bahasa Indonesia-Bahasa Jawa

Pembuatan *dataset* NLI lintas-bahasa untuk bahasa Indonesia-bahasa Jawa dibuat dengan menggunakan data dari IndoNLI (Mahendra et al., 2021). Kalimat hipotesis diterjemahkan ke dalam bahasa Jawa dengan menggunakan API ChatGPT OpenAI, sementara kalimat premis dibiarkan tetap dalam bahasa Indonesia. *Machine translation* (MT) dipilih untuk menerjemahkan kalimat hipotesis dari bahasa Indonesia ke bahasa Jawa karena cepat dan murah. ChatGPT OpenAI dipilih sebagai API penerjemahan karena terjemahan yang dihasilkan tidak bersifat

“*translationese*” (istilah yang dipergunakan oleh Budur et al. (2020) dalam penelitiannya untuk menyebut hasil terjemahan dengan machine translation, dikarenakan hasil terjemahan sangat kaku), serta hasil ChatGPT bersifat seperti percakapan. Bahasa Jawa yang dihasilkan oleh ChatGPT juga rata-rata menggunakan bahasa Jawa Ngoko. Untuk menjaga kualitas hasil terjemahan dan verifikasi apabila hasil terjemah mengubah label yang sudah ada, maka perlu dilakukan *quality assurance* oleh anotator.

### **III.2.1 Persyaratan Peserta Anotator**

Persyaratan peserta yang dapat berpartisipasi sebagai anotator yaitu:

1. Peserta harus memahami bahasa Indonesia secara baik dan benar.
2. Peserta harus dapat memahami bahasa Jawa secara baik dan benar.
3. Peserta merupakan warga asli suku Jawa dan dapat secara aktif berbahasa Jawa.

### **III.2.2 Metode Evaluasi *Dataset***

Berikut langkah-langkah evaluasi *dataset* yang dilakukan oleh anotator:

1. Setiap *set* data terdiri dari sebuah kalimat premis dalam bahasa Indonesia dan pasangan terjemahannya dalam bahasa Jawa, serta kalimat hipotesis dalam bahasa Indonesia dan pasangan terjemahannya dalam bahasa Jawa.
2. Setiap anotator akan diberikan *set* data yang sama.
3. Setiap anotator melakukan evaluasi terhadap *set* data yang diberikan dengan cara berikut:
  - a. Menilai kualitas terjemahan antara 1-5 (5 paling bagus).
  - b. Mengecek hasil terjemahan bebas dari kesalahan tata bahasa, ejaan, maupun tanda baca. Apabila terdapat kesalahan, maka anotator menandai data tersebut dengan memberikan nilai 1 pada kualitas terjemahan.
  - c. Perbaikan terhadap hasil terjemahan dilakukan oleh anotator independent sesuai dengan aturan pada bagian III.1.5.
  - d. Mengecek apabila hasil terjemahan mengubah hubungan semantik antar kalimat.

### III.2.3 Hubungan Semantik Antar Kalimat

Hubungan semantik antar pasangan kalimat premis dan hipotesis dapat dilihat berdasarkan deskripsi berikut:

1. Sepasang kalimat premis-hipotesis dapat dikatakan *entailment* apabila dapat disimpulkan bahwa kalimat hipotesis benar berdasarkan informasi yang ada pada kalimat premis.
2. Sepasang kalimat premis-hipotesis dapat dikatakan *contradiction* apabila dapat disimpulkan bahwa kalimat hipotesis salah berdasarkan informasi yang ada pada kalimat premis.
3. Sepasang kalimat premis-hipotesis dapat dikatakan *neutral* apabila berdasarkan informasi yang ada pada kalimat premis, kebenaran kalimat hipotesis tidak dapat ditentukan, atau tidak ada cukup informasi.

### III.2.4 Pemilihan Label Emas

Label emas (*gold label*) merupakan label utama yang akan dipakai untuk menandai hubungan semantik antara sebuah kalimat premis dengan pasangan kalimat hipotesisnya. Hasil dari pengecekan hubungan semantik dari setiap anotator pada bagian III.1.2.3.c akan dicek dengan cara sebagai berikut:

1. Apabila sepasang kalimat premis-hipotesis memiliki label yang sama dari semua anotator, maka label tersebut menjadi label emas.
2. Apabila terdapat anotator yang memberikan label tidak sama dengan anotator lainnya, maka dilakukan pemilihan label emas dengan cara melihat mayoritas label yang diberikan.
3. Apabila tidak ada mayoritas label yang dapat dipilih dari label-label yang diberikan oleh anotator, maka diberikan label ***broken***, dan pasangan kalimat tersebut dikeluarkan dari *dataset*.

### III.2.5 Perbaikan Teks Data

Teks hipotesis dalam bahasa Jawa yang dirasa oleh anotator tidak tepat, salah tata bahasa, ejaan, maupun tanda baca, yang ditandai dengan rata-rata nilai kualitas terjemahan kurang dari 3, dapat dilakukan perbaikan. Pembuatan kalimat baru dapat dilakukan dengan strategi berikut:

1. Penghilangan kata

Terdapat satu atau lebih kata yang dihapus dari kalimat premis.

2. Penambahan kata

Menambahkan satu atau lebih kata dari kalimat premis. Contohnya dapat menambahkan kata sifat, kata negasi, dan lainnya

3. Perubahan leksikal

Menggantikan satu atau lebih kata dari kalimat premis dengan sinonim, antonim, hipernim, ataupun hiponim.

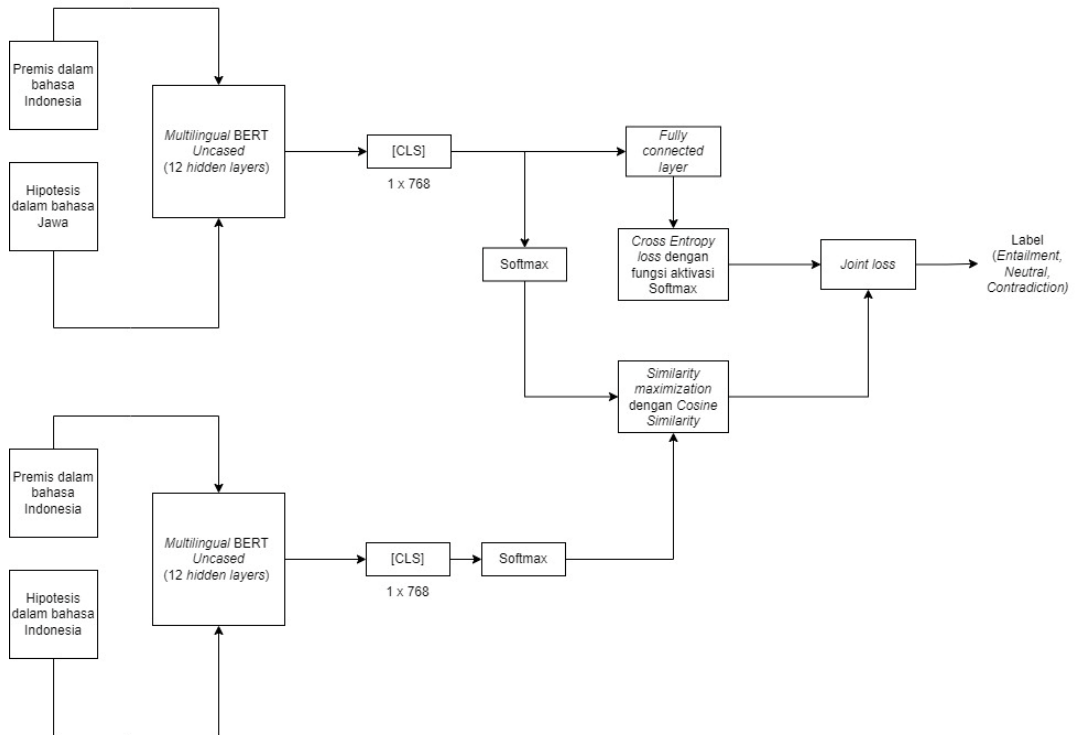
4. Parafrase

5. Perubahan struktur kalimat

Mengubah struktur kalimat, contohnya dari kalimat aktif menjadi kalimat pasif.

### III.3 Metode yang Diajukan

Untuk menangani permasalahan penalaran kalimat hipotesis yang kata kuncinya tersebar baik di awal, tengah, ataupun akhir kalimat premis, metode yang diajukan adalah menggunakan *transfer learning* yang diambil dari penelitian Bandyopadhyay et al. (2022) dengan perubahan pada perhitungan *similarity maximization* yang menggunakan *cosine similarity*. *Transfer learning* dilakukan dari model *baseline* yang sudah dilatih sebelumnya dalam bahasa Indonesia (model guru) ke model *baseline* yang menjadi tujuan dalam bahasa Indonesia-bahasa Jawa (model murid), seperti yang ditunjukkan pada Gambar III.1.



Gambar III.1 Skema metode transfer learning yang diajukan.

Pada metode yang diajukan, seperti yang ditunjukkan oleh Gambar III.1, model BERT *multilingual* mendapatkan masukan berupa pasangan kalimat premis-hipotesis. Selanjutnya, *output* dari model berupa *hidden state* terakhir diambil token pertamanya. Token pertama dari *hidden state* terakhir ini merupakan representasi dari token CLS. Vektor CLS ini dimasukkan ke *feed-forward layer*. Kemudian, probabilitas label dihitung dengan fungsi aktivasi Softmax. Perhitungan untuk memaksimalkan *similarity* dihitung dengan *cosine similarity* untuk menghasilkan *loss* gabungan antara model guru dengan model murid.

## DAFTAR PUSTAKA

- Arora, S., Liang, Y., & Ma, T. (2017). A Simple But Tough-To-Beat Baseline For Sentence Embeddings. *ICLR*, 1–16. <https://github.com/PrincetonML/SIF>
- Artetxe, M., Labaka, G., & Agirre, E. (2017). Learning bilingual word embeddings with (almost) no bilingual data. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 1, 451–462. <https://doi.org/10.18653/v1/P17-1042>
- Bandyopadhyay, D., De, A., Gain, B., Saikh, T., & Ekbal, A. (2022). A Deep Transfer Learning Method for Cross-Lingual Natural Language Inference. *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, 20–25. <https://github.com/google-research/>
- Bentivogli, L., Clark, P., Dagan, I., & Giampiccolo, D. (2010). The Sixth PASCAL Recognizing Textual Entailment Challenge. *Proceedings of the Third Text Analysis Conference (TAC 2010)*.
- Bentivogli, L., Clark, P., Dagan, I., & Giampiccolo, D. (2011). The Seventh PASCAL Recognizing Textual Entailment Challenge. *Proceedings of the Fourth Text Analysis Conference (TAC 2011)*.
- Bentivogli, L., Dagan, I., Dang, H. T., Giampiccolo, D., & Magnini, B. (2009). The Fifth PASCAL Recognizing Textual Entailment Challenge. *Proceedings of the Second Text Analysis Conference (TAC 2009)*, 15.
- Blacoe, W., & Lapata, M. (2012). A Comparison of Vector-based Representations for Semantic Composition. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 546–556.
- Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*, 4349–4357. <http://arxiv.org/abs/1607.06520>
- Budur, E., Özçelik, R., Güngör, T., & Potts, C. (2020). Data and Representation for Turkish Natural Language Inference. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8253–8267. <https://doi.org/10.18653/v1/2020.emnlp-main.662>
- Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St John, R., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil Google Mountain View, R. A. (2018). Universal Sentence Encoder for English. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174.
- Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised Learning of Universal Sentence Representations from Natural Language Inference



- Data. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 670–680. <http://arxiv.org/abs/1705.02364>
- Conneau, A., Rinott, R., Lample, G., Schwenk, H., Stoyanov, V., Williams, A., & Bowman, S. R. (2018). XNLI: Evaluating Cross-lingual Sentence Representations. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2475–2485.
- Dagan, I., Glickman, O., & Magnini, B. (2006). The PASCAL Recognising Textual Entailment Challenge. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3944 LNAI, 177–190. [https://doi.org/10.1007/11736790\\_9](https://doi.org/10.1007/11736790_9)
- Espana-Bonet, C., Varga, A. C., Barron-Cedeno, A., & Van Genabith, J. (2017). An Empirical Analysis of NMT-Derived Interlingual Embeddings and Their Use in Parallel Sentence Identification. *IEEE Journal on Selected Topics in Signal Processing*, 11(8), 1340–1350. <https://doi.org/10.1109/JSTSP.2017.2764273>
- Falke, T., Ribeiro, L. F. R., Ajie Utama, P., Dagan, I., & Gurevych, I. (2019). Ranking Generated Summaries by Correctness: An Interesting but Challenging Application for Natural Language Inference. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2214–2220. <https://doi.org/10.18653/v1/P19-1213>
- Faruqui, M., & Dyer, C. (2014). Improving Vector Space Word Representations Using Multilingual Correlation. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 462–471.
- Flek, L., McCrae, J., Sajjad, H., Piryani, R., Kann, K., Copyright, frai, Ruiz, M., Ebrahimi, A., Mager, M., Oncevay, A., Ortega, J. E., Rios, A., Fan, A., Gutierrez-Vasques, X., Chiruzzo, L., Giménez-Lugo, G. A., Ramos, R., Vladimir Meza Ruiz, I., Mager, E., ... Thang Vu, N. (2022). AmericasNLI: Machine translation and natural language inference systems for Indigenous languages of the Americas. *Frontiers in Artificial Intelligence*, 5. <https://doi.org/10.3389/frai.2022.995667>
- Giampiccolo, D., Dang, H. T., Magnini, B., Dagan, I., & Dolan, B. (2008). The Fourth PASCAL Recognizing Textual Entailment Challenge. *Proceedings of the First Text Analysis Conference (TAC 2008)*, 9.
- Gouws, S., Bengio, Y., & Corrado, G. (2015). BiBOWA: Fast Bilingual Distributed Representations without Word Alignments. *Proceedings of the 32 Nd International Conference on Machine Learning*, 37, 748–756. <http://arxiv.org/abs/1410.2455>
- Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., & Klakow, D. (2021). A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2545–2568. <https://doi.org/10.18653/v1/2021.naacl-main.201>

- Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification*. <http://arxiv.org/abs/1801.06146>
- Hu, H., Richardson, K., Xu, L., Li, L., Kuebler, S., & Moss, L. S. (2020). OCNLI: Original Chinese Natural Language Inference. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 3512–3526. <http://arxiv.org/abs/2010.05444>
- Jundi, I., & Lapesa, G. (2022). How to Translate Your Samples and Choose Your Shots - Analyzing Translate-train and Few-shot Cross-lingual Transfer. *Findings of the Association for Computational Linguistics: NAACL 2022*, 129–150.
- Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Torralba, A., Urtasun, R., & Fidler, S. (2015). Skip-Thought Vectors. *Advances in Neural Information Processing Systems*, 3294–3302. <http://arxiv.org/abs/1506.06726>
- Kryści, W., Nski, K., Mccann, B., Xiong, C., & Socher, R. (2020). Evaluating the Factual Consistency of Abstractive Text Summarization. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 9332–9346. <https://doi.org/10.18653/v1/2020.emnlp-main.750>
- Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2018). Visualizing and Understanding Neural Models in NLP. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2, 91–96. <http://arxiv.org/abs/1506.01066>
- Liu, B., Kang, H., Li, H., Hua, G., & Vasconcelos, N. (2020, May 27). Few-Shot Open-Set Recognition using Meta-Learning. *CVF Conference on Computer Vision and Pattern Recognition*. <http://arxiv.org/abs/2005.13713>
- Luong, M.-T., Pham, H., & Manning, C. D. (2015). Bilingual Word Representations with Monolingual Quality in Mind. *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 151–159.
- Maccartney, B., & Manning, C. D. (2008). Modeling Semantic Containment and Exclusion in Natural Language Inference. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 521–528.
- Mahendra, R., Aji, A. F., Louvan, S., Rahman, F., & Vania, C. (2021). *IndoNLI: A Natural Language Inference Dataset for Indonesian*. <https://doi.org/10.18653/v1/2021.emnlp-main.821>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). *Distributed Representations of Words and Phrases and their Compositionality*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b, January 16). Efficient Estimation of Word Representations in Vector Space. *Proceedings of the 1st International Conference on Learning Representations (ICLR 2013)*. <http://arxiv.org/abs/1301.3781>
- Mikolov, T., Le, Q. V., & Sutskever, I. (2013). *Exploiting Similarities among Languages for Machine Translation*. <http://arxiv.org/abs/1309.4168>

- Mogadala, A., & Rettinger, A. (2016). Bilingual Word Embeddings from Parallel and Non-parallel Corpora for Cross-Language Text Classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 692–702.
- Moritz Hermann, K., & Blunsom, P. (2014). Multilingual Models for Compositional Distributed Semantics. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 1, 58–68.
- Parikh, A. P., Täckström, O., Das, D., & Uszkoreit, J. (2016). A Decomposable Attention Model for Natural Language Inference. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2249–2255.  
<https://doi.org/10.18653/v1/D16-1244>
- Pfeiffer, J., Vulićević, I., Gurevych, I., & Ruder, S. (2020). MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7654–7673. <https://doi.org/10.18653/v1/2020.emnlp-main.617>
- Qi, K., Wan, H., Du, J., & Chen, H. (2022). *Enhancing Cross-lingual Natural Language Inference by Prompt-learning from Cross-lingual Templates* (Vol. 1). Long Papers.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992.  
<http://arxiv.org/abs/1908.10084>
- Schwenk, H., & Douze, M. (2017). Learning Joint Multilingual Sentence Representations with Neural Machine Translation. *Proceedings of the 2nd Workshop on Representation Learning for NLP*, 157–167.
- Søgaard, A., Zeljko Agić, Martínez Alonso, H., Plank, B., Bohnet, B., & Johannsen, A. (2015). Inverted indexing for cross-lingual NLP. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 1, 1713–1722.
- Soudani, H., Mojab, M. H., & Beigy, H. (2022). Persian Natural Language Inference: A Meta-learning approach. *Proceedings of the 29th International Conference on Computational Linguistics*, 4306–4319. <https://github.com/HassanMojab/MetaNLI>.
- Tandon, C., Bongale, P. J., Palivela, H., R, S. R., & R, N. C. (2021). UT-ATD: Universal Transformer for Anomalous Trajectory Detection by Embedding Trajectory Information. *Journal of Visual Language and Computing*, 2021(1), 27–32.  
<https://doi.org/10.18293/jvlc2021-n1-011>
- Trivedi, H., Kwon, H., Khot, T., Sabharwal, A., & Balasubramanian, N. (2019). Repurposing Entailment for Multi-Hop Question Answering Tasks. *Proceedings of the 2019 Conference of the North American Chapter of the Association for*

*Computational Linguistics: Human Language Technologies*, 2948–2958.  
<https://doi.org/10.18653/v1/N19-1302>

Upadhyay, S., Faruqui, M., Dyer, C., & Roth, D. (2016). Cross-lingual Models of Word Embeddings: An Empirical Comparison. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1661–1670.  
<https://doi.org/10.18653/v1/P16-1157>

Wang, X., Ruder, S., & Neubig, G. (2022). Expanding Pretrained Models to Thousands More Languages via Lexicon-based Adaptation. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1, 863–877. <https://doi.org/10.18653/v1/2022.acl-long.61>

Welleck, S., Weston, J., Szlam, A., & Cho, K. (2019). Dialogue Natural Language Inference. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3731–3741. <https://doi.org/10.18653/v1/P19-1363>

Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2016). *Towards Universal Paraphrastic Sentence Embeddings*. 1–14. <http://arxiv.org/abs/1511.08198>

Zhang, Y., Gaddy, D., Barzilay, R., & Jaakkola, T. (2016). Ten Pairs to Tag-Multilingual POS Tagging via Coarse Mapping between Embeddings. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1307–1317.

Zhou, X., Wan, X., & Xiao, J. (2016). Cross-Lingual Sentiment Classification with Bilingual Document Representation Learning. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1403–1412.