

# オープンドメインな情報提供のための zero-shot 学習に基づく自然言語理解

## Zero-shot Learning of Natural Language Understanding for Information Provision in Open Domains

貞光 九月\*<sup>1</sup> 本間 幸徳\*<sup>1</sup> 東中 竜一郎\*<sup>1</sup> 松尾 義博\*<sup>1</sup>  
Kugatsu Sadamitsu Yukinori Homma Ryuichiro Higashinaka Yoshihiro Matsuo

\*<sup>1</sup>NTT メディアインテリジェンス研究所  
NTT Media Intelligence Laboratories

We propose a new zero-shot learning method for natural language understanding that can take into account common question types across different domains as new features and can deal with sequential structures. In our experiments, our methods achieved higher accuracy between totally different domains (insurance and sightseeing) than baseline methods.

### 1. はじめに

近年の音声対話エージェントの登場により、自然文による情報提供を行うための自然言語理解技術に注目が集まっている。しかし、多様なドメインに対して自然言語理解技術を展開する際、ドメイン毎に学習用データを作成するには多大なコストを要する。このような課題を背景として、多くの zero-shot 学習や転移学習が提案されてきた [Yazdani and Henderson2015, Ferreira et al.2015, Kim et al.2015]。従来の zero-shot 学習では、自然文中の単語と知識ベース内のエントリの類似度を求める際に主に単語の分散表現 (word embedding) を用いることで、比較的近いドメイン間での転用を可能としている。しかし、ドメインが大きく異なる場合においては、単語の分散表現だけで十分な精度を維持することは困難である。また自然文中の各単語同士が、独立で知識ベースにマッピングされることから、自然文に含まれる文脈情報を十分に活かしてきれていないことも課題である。

本稿では上記の課題を解決する新たな zero-shot 学習を提案する。提案手法の特徴は、異なるドメイン間でも頑健に自然言語理解を行うために、ドメイン間で共通性を持つ汎用的な質問タイプを zero-shot 学習の新たな素性として導入する点、そして、知識ベースの構造と自然文中の系列に則した系列ラベリングを可能とした点、の 2 点である。提案手法を評価するために、異なる 2 つのドメイン (観光と保険ドメイン) 間で実験を行い、ベースラインと比較して精度を改善したことを示す。

### 2. 本稿で扱う自然言語理解タスク

図 1 に本稿で扱うタスクの概要を示す。元ドメイン (図中では“観光/sightseeing”) における訓練データは、人手で知識ベースとの対応付けをアノテートされた自然文集と、元ドメインの知識ベースであり、ここから自然言語理解モデルが学習される。一方、実行時 (評価時) に想定する入力、目的ドメイン (図中では“保険/insurance”) における 1 つの自然文と目的ドメインの知識ベースであり、出力は自然文と目的ドメイン知識ベース中のエントリとのマッピング結果である。このマッピングが正しく行えたか否かで評価を行う。

本稿で扱う知識ベースは、RDF に従って三つ組要素、subject, predicate, object (以下、*subj*, *pred*, *obj*) から構成される。図 1 と同じ知識ベースの例を三つ組形式で以下に示す。

- 観光 KB: {*subj* = “寿寺”, *pred* = “アクセス”, *obj* = “長谷駅から徒歩 7 分” }
- 保険 KB: {*subj* = “安心車保険”, *pred* = “保険料”, *obj* = “1000 円/月” }

以下では便宜的に、“*subj*”, “*pred*”, “*obj*” の三値のいずれかを取る仮想的なラベルをトリプルラベル  $c$  と呼び、“寿寺” や “長谷寺から徒歩 7 分” のように、自然言語によって記述されるトリプルラベルに対する具体的なエントリをトリプルエントリ  $g$  と呼ぶ。

### 3. 知識ベース構造と自然文系列を考慮した zero-shot 学習モデル

本節では、転移学習として提案された先行研究に基づき [Kim et al.2015]、知識ベースの構造と自然文の系列を考慮可能な zero-shot 学習モデルを提案する。Kim らは、自然文系列を考慮可能な Hidden Unit CRFs (HUCRF) を用いており、HUCRF における最終出力であるトリプルエントリ  $g$  へのマッピング確率は以下のように表せる。 $p(g|w) = \sum_{z \in \{0,1\}^n} P(y, z|x)$ , ここで  $p(y, z|x) \propto \exp(\theta^T \Phi(x, z) + \gamma \Psi(z, y))$ ,  $\theta, \gamma$  は図 2 中でそれぞれ下層, 上層の重みパラメータで、 $\Phi, \Psi$  はそれぞれ、下層, 上層の素性関数である。Kim らの転移学習は、 $z$  の自然文系列での出現傾向が、ドメイン間で類似するという仮説に基づき、パラメータ  $\theta$  をドメイン間で転移させる手法を提案した。

これに対し我々は、質問文中で用いられるトリプルラベル  $c$  の構造にも、ドメイン間で類似性があると考え、中間層である隠れ変数  $z$  からトリプルラベル  $c \in \{subj, pred, obj, NIL\}$  へ値域を拡張した。図 2 の下部に提案モデルを図示する。本モデルは二層のモデルから成り、下層において  $c$  を最尤系列としてデコーディングした後に、上層モデルの  $g$  へのマッピング処理を行うカスケード型のモデルとなっている。 $c$  はドメイン間共通で推定されるトリプルラベルであり、 $c$  によって規定されるトリプルエントリ候補集合 (=  $G_c$ ) の中からトリプルエントリ  $g$  が選択される。下層モデルにおいて、トリプルラベルの系列を自然文中から直接推定することで、従来の zero-shot 学習では取り扱われなかった知識ベース構造とテキスト構造

連絡先: 貞光 九月, NTT メディアインテリジェンス研究所, 神奈川県横須賀市光の丘 1-1, sadamitsu.kugatsu@lab.ntt.co.jp

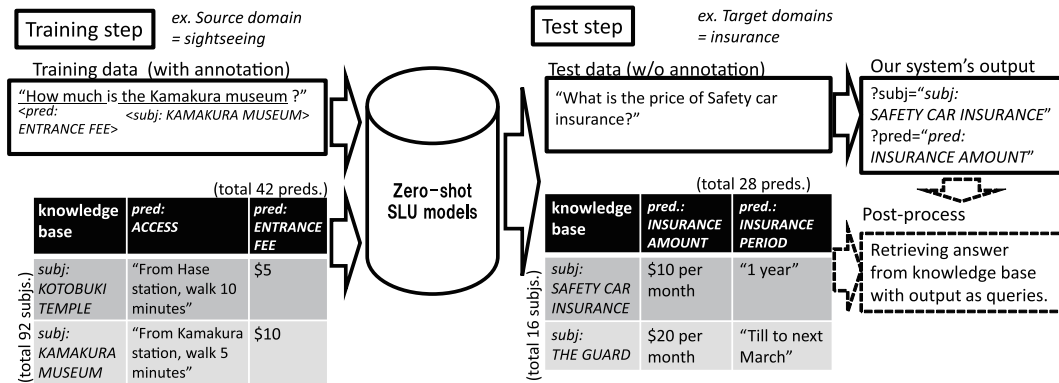


図 1: 本 zero-shot 学習タスクの概要 (Overview of our zero-shot learning task.)

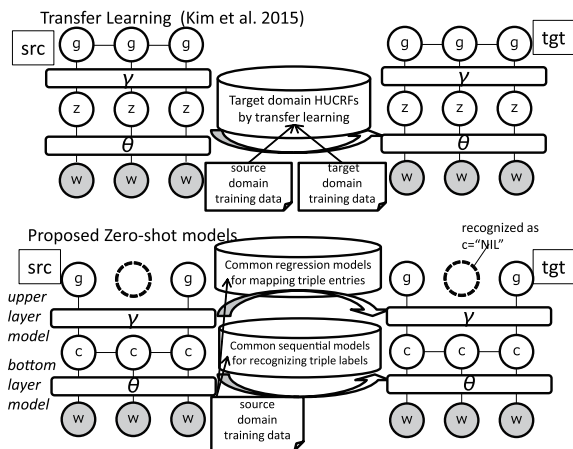


図 2: 提案システム (下段) と従来の転移学習 (上段) の比較 (Illustration of our system in comparison with early transfer learning.)

を取り入れ、一方上層モデルにおいて、ドメインが変化した場合でもトリプルエントリとのマッピングを可能としたことで、従来の転移学習に対して、zero-shot 学習が可能になるという利点が生じる。

以下、提案モデルの各層に分けて詳説する。図 2 の提案モデルの下層は、トリプルラベル  $c$  を推定するための系列モデルである。本系列モデル学習時には、元ドメイン側において、各単語のトリプルエントリ  $g$  へのアノテートラベルを、トリプルラベル  $c$  に一意に抽象化し、学習データとして用いる。例えば、 $\langle \text{pred:ENTRANCE FEE} \rangle$  がアノテートされている場合、 $\langle \text{pred} \rangle$  に抽象化する。これにより、ドメイン間でも分類対象を  $c$  に共通化することが可能となる。

系列モデルには CRF を用い、素性としては、広く用いられている単語表層、品詞、単語 ngram 素性等に加え、知識ベースとの類似度を取り扱うために、 $subj, pred, obj$  それぞれのトリプルエントリ集合  $G_c$  の中で単語分散表現 [Mikolov et al.2013] に基づく cosine 類似度が最大となるエントリ  $g$  との類似度を用いた。

次に図 2 の上層モデルにあたる、トリプルエントリ  $g$  にマッピングするための回帰モデルについて述べる。本モデルは、下層で推定されたトリプルラベル  $c$  を、複数の類似度を考慮することで、トリプルエントリ  $g$  にマッピングすることが目的となる。ここで複数の類似度を統合するのに際し、本稿では回帰

モデルを用いる [Fan et al.2008]。回帰モデルの学習においては、正しいマッピングに対し 1 の値を、同じトリプルラベルを持つトリプルエントリ候補  $G_c$  の中から生成される誤ったマッピングに対し 0 の値を付与することで、未知のドメインに対しても、値を付与し、トリプルエントリ候補のうち最大値を取るものをマッピング先と推定する。

$g_i = \arg \max_{g \in G_c} p(\gamma \cdot \Psi(g, c_i, w_i))$ , ここで  $\gamma$  は回帰モデルにおける各素性の重み、 $i$  は単語位置、 $\Psi$  は素性関数であり、素性関数の詳細について次節で述べる。

## 4. 知識ベースマッピングのための汎用的素性

### 4.1 基本素性

知識ベースのトリプルエントリ  $g$  へのマッピングのための回帰モデルにおいて、ドメイン間で汎用となるような素性 (類似度) の設計は重要である。我々はまず、先行研究に従い、単語  $w$  とトリプルエントリ  $g$  との表層類似度、文字の重複率、編集距離等の表層的類似度に加え、前節同様、単語の分散表現に基づく類似度を素性として用いる。単語の分散表現では、ドメイン依存型、ドメイン非依存型の 2 つのモデルを作成する [Yazdani and Henderson2015, Ferreira et al.2015, Mikolov et al.2013]\*<sup>1</sup>。具体的には、ドメイン依存型では元ドメイン及び目的ドメインの知識ベースに含まれるトリプルエントリを検索クエリとして収集した WEB テキスト 約 18.6 万文を用いてモデルを学習し、一方ドメイン非依存型では、Wikipedia 約 1200 万文を用いてモデルを学習した。

### 4.2 汎用質問タイプ

前節で挙げた基本素性は、テキスト中の文字列とトリプルエントリの類似度を測る上で強力であるが、ドメインが変化した場合に、正解のマッピングであるにも関わらず、表層情報や分散表現において類似しない場合がある。特に、 $pred$  は図 1 の例にも示す通り、表層を大きく変えて出現する「博物館はいくら?」という自然文の場合、単語「いくら」と知識ベース内のトリプルエントリ「入場料」を単語の分散表現に基づいて繋ぎ合わせることは困難である。

本課題を解決するため、我々はドメインに依存しない汎用質問タイプ (QT) を新たな素性として用いる。汎用質問タイプを導入するにあたっては、それがマッピングにおいて有益な情報とならなければならないのはもちろんのこと、新規のアノテーションコストをできるだけ低く抑えられる必要がある。そこで我々は、文の質問タイプと、 $pred$  に対応する質問タイ

\*1 <https://code.google.com/p/word2vec>

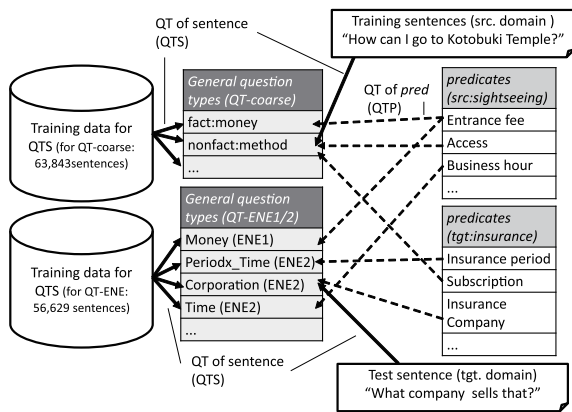


図 3: 提案手法における汎用質問タイプの利用 (The utilization of general question types)

ブの同一性に着目した。文の質問タイプについては、先行研究において質問応答システムにおいて効果が認められている [Ferruci et al.2012, Higashinaka et al.2013]。一方、*pred* は日本語文末に現れやすいことから、文の質問タイプ (以下 QTS) と、*pred* に対応する質問タイプ (以下 QTP) との間に、高い相関があるのではないかと考え、これを新しい素性として用いる。また本手法に必要なアノテーションコストとしては、ドメイン間で汎用的に使い回すことの可能な、文の質問タイプへの正解付与と、*pred* に対する質問タイプ付与のみであり、質問文に含まれる全トリプルエントリに対するマッピング正解付与は不要となるため、アノテーションコストを削減できる。

本稿では、Higashinaka et al., 2013 の研究を参考に、2 種類の汎用質問タイプを用いる。1 つ目は non-factoid な質問に対する分類を含む 23 種の粗い分類体系であり、“QT coarse”と呼ぶ。今回扱う知識ベースには、“アクセス手段”を説明するような、non-factoid 型のリテラル値を含んでいるために、これらの質問を分類可能な本体系を用いた (図 1 の知識ベース例参照)。2 つ目は factoid 型質問に対する詳細分類であり、関根の拡張固有表現体系に基づいた分類体系である [Sekine2008]。本体系は最大三階層から構成されるが、タイプの共起や一致を手掛かりとする際に分類が細かすぎると共起素性としてスパースになる問題が生じるため、第一階層 (28 種) および第二階層 (87 種) のみ用いることとした。

提案手法における汎用質問タイプの適用方法を図 3 に示す。まず、文と *pred* それぞれに対し、汎用質問タイプとの対応である QTS, QTP を取得する (図中の実線・破線矢印)。文の QTS を推定する際には、ロジスティック回帰による多値分類器を用い [Fan et al.2008]、素性には表層素性、ngram 素性等一般的に用いられる素性を用いる。学習用質問文データ作成においては、汎用質問タイプを網羅するように作成しており、本分類器を学習したデータ量、および二交差検定によって精度評価した結果、“QT-coarse”、“QT-ENE1” (1 階層目)、“QT-ENE2” (2 階層目) について、それぞれ 86.94%、86.01%、78.61% の精度を示すことを確認している。一方 QTP に対しては、元ドメイン、目的ドメイン双方の全 *pred* に対し、汎用質問タイプのタイプ名との類似度を、単語分散表現の cosine 尺度により得た後、人手で拡張・修正を行った [Mikolov et al.2013]。

次に QTS と QTP の汎用質問タイプの素性としての使用方法について述べる。まず、下層系列モデルにおいて入力自然文中 *pred* が文末位置で認定された場合、入力文に対し、先に述べた汎用質問タイプ分類器によって QTS を推定し、そのう

ち上位  $N$  個の汎用質問タイプを得る (本稿では  $N = 3$  に設定)。続く上層回帰モデルにおいて、各 *pred* 候補に対し個別に素性化の際、QTS と QTP の共起に関して本稿では以下の 2 種類の素性を導入する。一つ目はマッチ素性 (“q-match”) で、QTS と QTP が一致しているか否かを示す素性である。二つ目はペア素性 (“q-pair”) で、QTS と QTP を直接ペアとして素性化するものである。前者は目的ドメインにおける汎用質問タイプが、必ずしも元ドメインで現れる必要がないため、ドメインの変化に対しても頑健に働くと期待される。後者はより細かく情報量をモデルに与えることができるため、目的ドメインと元ドメインで共通する汎用質問タイプが存在する場合には、精度向上が期待される。また、各素性の素性値は、質問文分類器が出力する信頼度を用いた。例えば、“q-match”素性を用いる際、推定された QTS と現在候補としている *pred* の QTP が一致している場合、その素性値は当該 QTS の質問文分類器が出力する信頼度とした。

一方、元ドメインのトリプルエントリと目的ドメイン間で似ているトリプルエントリを直接リンクし、素性化する方法も考えられる [Li et al.2011]。しかし、元ドメインと目的ドメインが大きく異なる場合、それらのトリプルエントリが重複するとは限らず、網羅性に問題がある。例えば図 3 では “Corporation” に相当するエントリが元ドメインである “観光” には存在していない場合を示している。これに対し、汎用質問タイプを用いることで、幅広いドメインに対応することが可能となる。

## 5. 実験と考察

### 5.1 実験設定

本節では 2 つの全く異なるドメインを評価対象とすることで、提案モデルの効果を確認した。用いたドメインは “観光” と “保険” ドメインであり、文の数と各トリプルレベルの統計量は以下の通りである。

- 観光: 937 文, *subj* 825 か所, *pred* 868 か所, *obj* 371 か所
- 保険: 443 文, *subj* 289 か所, *pred* 393 か所, *obj* 368 か所

ベースラインとして、系列モデルからリンク素性を取り除き、ユニグラム素性である単語分散表現素性等の基本素性のみを用いることで、先行研究 [Yazdani and Henderson2015, Ferreira et al.2015] を模した実験条件とした。汎用質問タイプについては、使用しない場合 (“q-none”)、“q-match”のみを使った場合、“q-pair”も併用した場合 (“q-match + q-pair”) の 3 通りを試した。

回帰モデル導入の欠点として、各トリプルエントリの回帰スコア算出時に、他のトリプルエントリ候補の素性値との直接的な比較が行えない点が挙げられる。そこで、エントリ同士の素性を相対的な値として用いるために、当該素性のうち、トリプルエントリ候補集合  $G_c$  の中で最大となる素性値を、全トリプルエントリの素性として追加した場合 (w/ rel) についても実験を行った。

### 5.2 実験結果と分析

質問文からトリプルエントリへのマッピング結果に対し、適合率、再現率、F 値により比較を行った。チャンキング範囲は後続の質問応答において影響しないため評価に含めていない。*pred* に関する結果を表 1 に示す。

まず、汎用質問タイプ素性を加えることにより、全体的に精度が向上していることが確認できる。この時、“q-match+q-pair” と “QT none” の結果はマクネマー検定で  $p = 0.01$  で有意に差がであった。相対化素性 (“rel”) は「保険 観光」への適用においては効果があるものの、逆向きの適用においては

	src:insurance to tgt:sightseeing									src:sightseeing to tgt:insurance								
	QT none (baseline)			q-match (proposed)			q-match+q-pair proposed			QT none (baseline)			q-match (proposed)			q-match+q-pair (proposed)		
	pre	rec	F	pre	rec	F	pre	rec	F	pre	rec	F	pre	rec	F	pre	rec	F
Baseline (w/o link)	44.5	35.7	39.6	47.3	37.7	42.0	48.9	38.7	43.2	21.5	58.5	31.4	23.4	61.9	33.9	23.8	63.5	34.7
proposed (w/o rel)	50.9	26.1	34.5	61.1	32.5	42.5	62.6	33.9	44.0	42.9	54.0	47.8	42.7	54.3	47.8	45.5	57.8	50.9
proposed (w/ rel)	59.6	33.1	42.6	61.4	34.1	43.9	63.1	35.2	45.2	41.9	53.1	46.9	42.3	53.8	47.3	42.7	54.3	47.8
gold-seq.(w/o rel)	48.1	42.9	45.3	54.9	51.5	53.1	57.5	55.0	56.2	60.3	59.8	60.1	63.9	63.4	63.6	65.6	65.2	65.4
gold-seq.(w/rel)	58.5	57.3	57.9	58.9	57.6	58.3	60.7	59.5	60.1	61.6	61.1	61.3	65.1	64.6	64.9	64.6	64.1	64.4
in domain	src:sightseeing to tgt:sightseeing									src:insurance to tgt:insurance								
	55.5	54.9	55.2	56.3	55.3	55.8	59.9	58.8	59.3	58.9	54.3	56.5	59.9	55.1	57.4	60.5	55.8	58.0

表 1: 実験結果。下線部は各条件における最高精度を示す。ベースライン (w/o link features) および indomain については相対素性の有無において精度の良い方を示した。(Experimental results.)

Feat. type	w2v (domain)		q-match	q-match	q-match
	Surface	dependent	QT coarse	QT ENE1	QT ENE2
Normal	2.08	3.06	1.57	0.569	1.04
Rel.	-0.249	-1.26	-0.290	-0.484	-0.727

表 2: 重みパラメータの学習結果 (The weights of features.)

むしろ悪化するという結果であった。これは、観光ドメインにおいては、相対化素性を用いずとも、正解となるトリプルエントリに対する素性値が、他のトリプルエントリと比べて十分に特徴的であったためと考えられる。その他の汎用質問タイプに依存しない *subj* と *obj* については以下の F 値を示しており、おおそ系列モデルを用いたことで改善傾向を示した。

- “観光 保険”: (proposed) *subj*=74.2, *obj*=63.7 : (baseline) *subj*=67.3, *obj*=61.7
- “保険 観光”: (proposed) *subj*=72.9, *obj*=32.4 : (baseline) *subj*=76.3, *obj*=31.3

表 2 は回帰モデル訓練後の各素性に対する重みである (相対素性も併記)。なお、素性値側の値は全て 0 から 1 の間に正規化されている。これにより、QT coarse の q-match 素性が比較的効果的であったことが確認できる。

さらに分析のため、系列モデルが全て正しいトリプルラベルを出力した場合を仮定した実験を行った。表 1 において “gold-seq” と付記したものが、その結果である。当然ながら、自動推定の場合に比べ結果が改善しているものの、それでも改善の余地は大きいことが確認できる。この結果により、後段のマッピングモデルの精度向上の方が課題として大きいことが示唆された。また、本実験においても汎用質問タイプ素性の効果は持続していることが確認でき、前段の系列モデル精度が向上した場合でも、引き続き QT の効果が期待される。

最後に、zero-shot ではなく、同ドメイン内で学習し、10 交差検定により評価した場合の結果を表中の “indomain” に示す。この結果から、zero-shot 学習と同ドメイン学習の間には依然隔たり (45.2 と 59.3, 47.8 と 58.0) があり、この差分を解消していくことが今後の課題となる。

## 6. 結論

本稿では、異なるドメイン間での zero-shot 学習の精度を改善するため、ドメイン間に共通な特徴として、汎用質問タイプ素性と知識ベース構造に基づく系列モデルを用いる手法を提案し、実験においてそれらが有効であることを確認した。

本稿では学習ドメインは単独として扱ったが、今後は複数のドメインを学習に用いることで、その中でドメイン間で共通する素性を導出し、精度改善へと結び付けたい。

## 参考文献

- [Fan et al.2008] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.
- [Ferreira et al.2015] Emmanuel Ferreira, Bassam Jabaian, and Fabrice Lefevre. 2015. Online Adaptive Zero-Shot Learning Spoken Language Understanding Using Word-Embedding. In *Proceedings of the ICASSP*, pages 5321–5325.
- [Ferruci et al.2012] David Ferruci, Anthony Levas, Sugato Bagchi, David Gondek, and Erik Mueller. 2012. Watson: Beyond Jeopardy! *Artificial Intelligence*, 199:93–105.
- [Higashinaka et al.2013] Ryuichiro Higashinaka, Kugatsu Sadamitsu, Kuniko Saito, and Nozomi Kobayashi. 2013. Question answering technology for pinpointing answers to a wide range of questions. *NTT Technical Review*, 11(7).
- [Kim et al.2015] Young-Bum Kim, Karl Stratos, Ruhi Sarikaya, and Minwoo Jeong. 2015. New Transfer Learning Techniques for Disparate Label Sets. In *Proceedings of the ACL/IJCNLP*, pages 473–482.
- [Li et al.2011] Xiao Li, Ye Yi Wang, and Gokhan Tur. 2011. Multi-task learning for spoken language understanding with shared slots. In *Proceedings of the INTERSPEECH*, pages 701–704.
- [Mikolov et al.2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- [Sekine2008] Satoshi Sekine. 2008. Extended named entity ontology with attribute information. In *Proceedings of the LREC*.
- [Yazdani and Henderson2015] Majid Yazdani and James Henderson. 2015. A Model of Zero-Shot Learning of Spoken Language Understanding. In *Proceedings of the EMNLP*, pages 244–249.