

# 自動意味役割付与のための役割集合の設計

松林 優一郎

東京大学

辻井 潤一

東京大学, マンチェスター大学, NaCTeM

## 1 はじめに

近年、PropBank (Kingsbury と Palmer, 2002) や FrameNet (Ruppenhofer ら, 2006) 等の意味役割付与コーパスの登場と共に、機械学習による意味役割の自動付与が盛んに研究されてきた。意味役割に代表される文の意味構造を解析する技術は、質問応答システムや機械翻訳、言い換えなどの自然言語処理応用技術に有用であると考えられている。

これら代表的な二つの意味役割付与コーパスは、文中の単語(主に述語)がフレームと呼ばれる特定の項構造を持つという考えに基づく。例えば述語 buy.v はそれぞれ図 1 のようなフレームを持つと定義されている。

buy.v	FrameNet	PropBank
フレーム	Commerce.buy	buy.01
意味役割	Buyer Goods Seller Money Recipient ...	A0: buyer A1: thing bought A2: seller A3: paid A4: benefactive

図 1 PropBank と FrameNet のフレーム比較

各フレームの構成要素は意味役割と呼ばれ、文中の特定の句で埋められる。役割は基本的にフレームによって定義されるため、役割の数は膨大である。実際に、PropBank は単語の意味ごとにフレームを定義しているためフレーム数は 3,323 に及び、FrameNet では、単語間でフレームを共有しているためフレーム数は 795 であるが、コーパス中に実例として出現する役割の数は 4,589 とやはり膨大である。このことは結果として、学習の際の各役割の実例不足を招く。PropBank の各フレームあたりの実例数は約 12 であり、FrameNet では役割の約半数が 10 個以下の実例しか持たない。

このような問題が起こるのは、本来異なるフレーム間で共有されるべき、それぞれの役割に類似の性質(統語的性質、役割を埋める語彙の傾向など)を、全く別のものとして扱うためである。既存研究の幾つかは、この疎データ問題を緩和するため、役割を互いに交わらない粗いクラスで置き換え、その新たなクラスを単位として分類を行った。しかしそのような方法では、個々の役割が持つ多様な性質の、ある一側面を纏め上げることしか出来ない。

我々のここでの仮定は、個々の役割は幾つかの異なる性質の組み合わせで表現されることで、より正しく分類される、というものである。そこで我々は、役割間の共通性を表す異なる情報を元に四種類の役割集合を設計し、それらを同時に学習で取り扱うための素性関数を設計する。今回、我々の実験には FrameNet コーパスを用いる。これは、FrameNet が役割横断的に付けられた豊富な情報を含むからである。実験では、意味役割付与問題のうち特に役割の疎データ問題と関係の深い役割分類と呼ばれる部分問題についてその精度を検証し、それぞれの役割集合が実例の少ない役割の分類に貢献することと、異なる種類の役割集合を混合するモデルが、意味役割分類の精度を向上させることを示す。

## 2 関連研究

PropBank における意味役割分類では、既存の殆どのシステムにおいて、フレームによる役割の違いは区別されず、形式的な役割の番号(A0-A5)とフレーム非依存の修飾句的役割(AM-\*)を正解の単位として役割分類が行われてきた。これらの形式的番号は A0 が動作主、A1 が非動作主、などの比較的緩い規則があるものの、同一番号の間で一貫した性質があるとは認め難い。Moschitti ら (2005) は PropBank の形式的ラベルを、さらにラベルの機能の違いごとに中心項、付加詞、連続項、共参照項の四つに分け、二段階の分類を行った。しかしこの方法も最終的には PropBank の形式的ラベルによる集合に結果が依存してしまう。Zapirain ら (2008) は PropBank の役割を VerbNet の主題役割に置き換えて、PropBank の形式的ラベルで置き換えた場合と精度を比較した結果、VerbNet の主題役割のほうがやや低い精度を示すと報告した。

FrameNet による意味役割分類では、Gildea と Jurafsky (2002) が彼らの定義した 18 個の主題役割へ役割を人手で置き換えることで分類精度が向上することを示した。また、Giuglea と Moschitti (2006)、Shi と Mihalcea (2005) は FrameNet の役割を VerbNet の主題役割へ対応付けたが、これらによる役割分類への寄与は直接検証されていない。Baldewein ら (2004) は個々のフレームに専用の分類器を作り、類似する役割の実例を使い回す方法を取った。彼らの方法もまた、複数の役割の一つの役割とみなして分類する方法であった。

## 3 役割分類

意味役割付与は複数の問題が絡み合った複雑なタスクであるため、これをフレーム想起単語特定(フレームを想起する単語の特定)、フレーム曖昧性解消(想起単語が取り得るフレームのうち正しいものを選択)、役割句特定(意味役割を持つ句の特定)、役割分類(役割句に正しい役割を割り当てる)、といった四つの部分問題に分けて解かれることが多い。今回我々は、これらの部分問題のうち、役割の疎データ問題が特に問題となる役割分類について、我々の手法の有効性を確かめる。

ここでは、既存研究に沿って以下のように役割分類の定義を与える。

- 役割分類とは、文、フレーム想起単語、フレーム、役割の候補、役割を持つ句が与えられたとき、それぞれの役割句に対して正しい役割を割り当てることである。

INPUT:
frame = GETTING
candidate roles = {Recipient, Theme, Reason, Time, Place, etc.}
sentence = It was in [a little book] [I] [once] [got GETTING] [for Christmas] [called Ghosts and Other Weird Creatures].
OUTPUT:
sentence = It was in [a little book <sub>Theme</sub> ] [I <sub>Recipient</sub> ] [once <sub>Time</sub> ] [got GETTING] [for Christmas <sub>Reason</sub> ] [called Ghosts and Other Weird Creatures <sub>Theme</sub> ].

図 2 役割分類における入出力の例

図2に具体的な入出力の例を示す。役割の候補はフレームによって与えられる。従来法である役割ラベルの置き換えは、これらの候補を全フレームに共通な粗い役割クラスに置き換えることに相当する。

## 4 手法

### 4.1 役割分類モデル

我々は役割分類問題を、可能な役割候補の中から正しい役割を選ぶ分類問題として定式化する。文中のある役割句が与えられたとき、その役割句に対する観測ベクトルを  $\mathbf{o}$  とする。また、この役割句に付与可能な役割候補を  $C(\mathbf{o})$  と表す。このとき、 $C(\mathbf{o})$  中の役割  $y$  が条件  $\mathbf{o}$  の下に観測される確率を以下のように置く。

$$p(y|\mathbf{o}) = \frac{1}{Z(\mathbf{o})} \exp\left(\sum_{i=1}^n \lambda_i f_i(y, \mathbf{o})\right) \quad (1)$$

$Z(\mathbf{o})$  は正則化項、 $F = \{f_1, \dots, f_n\}$  は  $n$  個の素性関数、 $\Lambda = \{\lambda_1, \dots, \lambda_n\}$  は  $F$  に関する重みベクトルを表す。我々は  $C(\mathbf{o})$  の役割の中から最尤な役割  $\tilde{y}$  を選択することで分類を行う。

$$\tilde{y} = \operatorname{argmax}_{y \in C(\mathbf{o})} \sum_{i=1}^n \lambda_i f_i(y, \mathbf{o}) \quad (2)$$

この際、個別のフレームの為に複数の分類器を作ることせず、全てのフレームで用いる一つのカテゴリ器を設計する。これは異なるフレームを横断して役割の共通性を学習するためである。素性関数の最適重み  $\Lambda$  は最大事後確率 (MAP) 推定によって求められる。我々は Limited-memory BFGS (L-BFGS) 法 (Nocedal, 1980) を用いて学習データの対数尤度を  $L_2$  正規化により最大化する。

### 4.2 役割の共通性の分類器への導入

一般的に、素性関数  $f_i$  はラベル  $y$  と観測  $\mathbf{o}_j$  の指示関数として定義される。例えば、主辞が *she* である句  $x$  が文中で意味役割 *Seller* を演じるという事象は以下の指示関数によって表される。

$$f_i(y, \mathbf{o}) = \begin{cases} 1 & (hw(x) = she \wedge y = Seller) \\ 0 & (otherwise) \end{cases} \quad (3)$$

役割の共通性を扱う為の従来の方法は、役割を別のラベルに置き換えることであった。例えば、役割 *Seller* がそのフレームで通常動作主として働くならば、*Seller* はラベル *Agent* として置き換え可能である。この操作は確かに役割の一側面 (動作主) を捉えているものの、役割の予測に有効なその他の手掛かりを取り逃すかもしれない。例えば、FrameNet の役割階層には、*Seller* は役割 *Donor* の性質を継承しているという情報がある。

このような役割に関する様々な観点を取り込む為に、役割の集合に対して発火する素性関数を設計する。例えば、役割集合 *Donor* に対する素性関数である

$$f_i(y, \mathbf{o}) = \begin{cases} 1 & (hw(x) = she \wedge y \in Donor) \\ 0 & (otherwise) \end{cases} \quad (4)$$

は、与えられた役割  $y$  が役割集合に属しているかを検定する。同様に、ラベル *Agent* に関しても、役割集合 *Agent* を考え、 $y \in Agent$  を検定する指示関数を設計することで、これらを同時に取り扱うことが出来る。

最大エントロピーモデルの利点は、ラベル候補に依存しない幅広い素性を取り扱える能力である。結果として、我々は様々な役割集合を素性関数として含むことが可能となる。この手法は従来のラベルの置き換えによる手法をより一般的な枠組みにするだけでなく、それぞれの役割集合の有用性を事前に決めることなく、学習の過程で素性の重みとして決定することが出来る。

### 4.3 単語列ベースの分類による構文解析エラーの緩和

幾つかの既存研究により、代替木によって構文解析エラーの影響を緩和する手法が報告された (Carreras と Marquez, 2005)。構文解析エラーが特に問題になるのは、正解の意味役割領域が構文木の構成素として存在しない場合である。それは、構文木の構成素を役割句の候補とする従来の多くのシステムが、役割を正しい領域に割り当てることが出来ないからである。そこで、我々は従来法の代わりに、単に単語列を役割句の候補にする方法を取る。この方法では、システムはある構文解析器による構文木の構成素領域だけでなく、文中の任意の単語列を役割句として考えることが出来る。構文木情報は役割句の候補領域を制限するものとしてではなく、その単語列が統語句と判別される確率と共に発火する学習器の素性として扱われる。このモデルでは、 $n$ -best 木を生成確率で重み付けされた素性として扱うことによって、 $n$  個の木全体を通してラベルの確率が最適化される。

## 5 役割集合の設計

役割の集合を設計する際には、それらの集合が効果的に役割の性質の共通点と相違点を捉えていることが重要である。幸い、FrameNet には各役割の性質を説明する幾つかの付加情報が付与されている。我々はこのうち、以下の三つの異なる観点からなる情報に特に注目し、役割集合の設計法を提案する。また、SemLink の画像情報を用いた VerbNet の主題役割による集合も検討する。以下では、それぞれの情報による集合の設計法について詳しく述べる。

### 5.1 役割名

FrameNet の意味役割は、各フレームに固有の役割として識別子が付けられているが、実際には、それらには人間に意味解釈可能な短い名前 (*Buyer*, *Seller* など) が付いている。これらの役割名は、ある程度体系立てて付けられており、その中には多数のフレームで共通して使われている名前も多々ある。これらの複数フレームに渡って同一名を持つ意味役割は厳密には違う識別子であるが、我々はこの、専門家が同一の名前を選んで定義した役割には、何らかの共通な性質があるという仮定を置く。この仮定に基づいて、例えば *Commerce\_buy* フレームの *Buyer* と *Commerce\_sell* フレームの *Buyer*、或いは多くのフレームで出現する *Time* という名の役割などを、同一集合の要素として纏め上げる。

### 5.2 役割間の階層関係

FrameNet のフレームには階層関係が定義されており、これらはフレーム間の七種類の有向関係 (継承、使用、起動、原因、観点、部分、先行) として記述されている。また、各フレームの意味役割の幾つかは、フレーム間関係を通して、他フレームの役割と一対一の役割間関係が定義されている。我々は、この情報を元にして、階層的な役割集合を設計する。基本的なアイデアは、下位概念にあたる役割は、その上位概念にあたる役割の性質を引き継いでいる、という仮定である。例えば、*Commerce\_buy* フレームの役割 *Buyer* は *Getting* フレームの役割 *Recipient* の意味を継承しており、また、*Killing* フレームの *Victim* と *Death* フレームの *Protagonist* は”死ぬも

の”という個体の性質を持っている。

これらの役割は共通の性質を持ちながらも、明確な意味の違いから異なる語彙で説明される。従って、これらの役割間の共通性は役割名による集合では捉えられない。役割間関係による集合は、特にこのような役割の性質を捉えることを目的とする。また、階層的な集合を通して最適な粒度の役割集合を探索することも目的とする。

集合は以下のように設計する。任意の役割  $r$  について、(1)  $r$  だけを要素として含む役割集合を作り、(2)  $r$  の先祖にあたる役割の集合に、 $r$  を追加する。起動型と原因型の関係については、下位フレームのほうがより中立的な意味や、結果の状態を表すため、これらの場合では関係の向きを逆転させる。また、先行型は事象の順序関係を表す関係であり、性質の包含関係が判断できないため、両方の方向を試す。

### 5.3 意味型

FrameNet では、多くの役割について意味型と呼ばれる型が付与されている。これは、意味役割を埋める句についての一選択制限である。例えば、*Self\_motion* フレームの役割 *Area* は意味型が *Location* で、場所を意味する句で埋められるべきだとしている。この情報は役割の粗いカテゴリを示唆しているため、同一の意味型を持つ役割を役割集合として取り扱う。

### 5.4 VerbNet の主題役割

VerbNet の主題役割は、述語の項に付けられた 23 種類の粗い意味分類である。主題役割は役割の機能的側面に特化しているため、これまで作成した集合に加えて、新たな観点として期待できる。我々は SemLink プロジェクト<sup>\*1</sup>によって作成された、FrameNet の意味役割と VerbNet の主題役割の間の写像を利用して、FrameNet の役割を VerbNet の主題役割による役割集合として纏め上げる。

### 5.5 既存手法との比較

FrameNet の役割名や階層関係を利用する方法は、Baldewein ら (2004) の研究に見られる。彼らは各フレームごとに専用の役割分類器を作り、学習データを増やす目的で、ある指標で類似であると判断された役割  $A$  と  $B$  について、 $A$  のための訓練例を、 $B$  のための訓練に再利用した。このときの指標に、役割名的一致や役割階層の上下関係を用いている。しかし、 $B$  のための訓練例が十分であるとき、この操作は不必要、あるいは悪い場合では  $A$  の実例が  $B$  の曖昧さを増加し、分類精度を落としてしまう。そのため、彼らはこの悪影響を懸念して、害をなす可能性のある素性を事前に取り除いたり、中心性<sup>\*2</sup>の低い役割のみ訓練例の再利用を行う工夫をしている。

一方で、我々は素性関数という形を介して、個別の役割と役割集合の双方を学習に用いてそれぞれに重みを最適化するので、この問題を考慮する必要が無い。従って、特にこれらの制限を設けず、全ての中心性の型について役割名による集合を作成し、全ての素性をを用いて学習する。主題役割を用いた例では Gildea と Jurafsky (2002)、Giuglea と Moschitti (2006) が人手で全ての役割を主題役割に写像したのに対し、SemLink は一部の役割 (主に中心性が *core* の役割) を写像するに止まるため、部分的な効果であることが予想される。

<sup>\*1</sup> <http://verbs.colorado.edu/semlink/>

<sup>\*2</sup> FrameNet では、そのフレームにおける各役割の中心性を四つの型 (*core*, *core-unexpressed*, *peripheral*, *extra-thematic*) に分けている。

表 1 各役割集合による分類精度

Feature	Micro	Macro
Baseline	89.00	68.50
+ 役割名	90.77	<b>76.58</b>
役割名で置換	90.23	76.19
+ 階層関係	90.25	72.41
+ 意味型	90.36	74.51
+ 主題役割	89.50	69.21
Total	<b>91.10</b>	75.92

表 2 役割の中心性ごとの役割名集合の効果

中心性	Micro
Baseline	89.00
Core	89.51
Peripheral	<b>90.12</b>
Extra-thematic	89.09
All	90.77

## 6 実験と議論

### 6.1 実験設定

実験には Semeval-2007 Shared task (Baker ら, 2007) の訓練データを用いる。このデータは FrameNet 第 1.3 版と追加の全文書役割付与データから成る。このうちランダムに抜き出した 10% をテストデータとして用い、残りのデータで 10 分割交叉検定を行い分類器を開発する。評価は各役割に対する Micro F1 値と Macro F1 値で行う。

### 6.2 ベースライン分類器

ベースラインには、Gildea と Jurafsky (2002)、Florian ら (2003)、Surdeanu ら (2003)、Xue と Palmer (2004)、Bejan ら (2004)、Pradhan ら (2005) によって有効と示された素性を使った分類器を用いる。用いた素性は、フレーム、フレーム想起単語、主辞、内容語、先頭/末尾単語、左右の兄弟ノードの主辞、句の統語範疇、句の位置、態、統語パス (有効/無向/部分)、支配範疇、主辞の Supersense、想起単語と主辞の組、想起単語と統語範疇の組、態と句の位置の組、等である。単語素性は品詞と語幹素性も用いている。構文解析には Charniak と Johnson (2005) の reranking parser を用い、従来の固有表現素性の代わりとして、Ciaramita と Altun (2006) の Super Sense Tagger を用いる。ベースライン分類器では役割集合を用いず、1-best 木で 88.34%、50-best 木で 89.00% の Micro F1 値を得た。

### 6.3 実験結果外観

表 1 には、それぞれの役割集合を用いたときの役割分類の Micro F1 と Macro F1 を示す。各タイプの役割集合でそれぞれ有意な精度向上が見られた。特に役割名による集合で顕著な向上を見ることが出来るため、これを細かく分析し、結果を表 2 と表 3 に示した。表 2 は、役割の中心性ごとにその型の役割だけを使って集合を作成した場合の性能を比較したものである。役割名集合が特に周縁的な役割を纏め上げるのに特に有効であることが分かる。これは表 3 に見られる様に、フレームに意味の依存が薄い役割がフレーム横断的に同一の役割名が付けられやすいという傾向によるものである。周縁的役割に分類される 1924 の役割は 250 という比較的小さな集合に纏まっている。主題役割による結果は、纏め上げられ

表 3 各中心性における役割及び役割集合の数と実例数

中心性	役割数	実例数/役割数	集合数	実例数/役割数
Core	1902	122.06	655	354.4
Peripheral	<b>1924</b>	<b>25.24</b>	<b>250</b>	<b>194.3</b>
Extra-thematic	763	13.90	171	62.02

表 4 実例の少ない役割に対する役割集合の効果

素性	実例数	Pre.	Rec.	Micro
baseline	≤ 10	63.89	38.00	47.66
	≤ 20	69.01	51.26	58.83
	≤ 50	75.84	65.85	70.50
+ 役割集合	≤ 10	72.57	55.85	63.12
	≤ 20	76.30	65.41	70.43
	≤ 50	80.86	74.59	77.60

表 5 役割間関係の種類による階層関係集合の精度比較

No.	関係	Micro
-	baseline	89.00
1	+ 継承 (親)	89.52
2	+ 継承 (祖先)	<b>89.70</b>
3	+ 使用 (親)	89.35
4	+ 使用 (先祖)	<b>89.37</b>
5	+ 観点 (親)	89.01
6	+ 部分 (親)	89.04
7	+ 部分 (先祖)	89.05
8	+ 原因 (子)	89.03
9	+ 起因 (子)	89.02
10	+ 先行 (親)	89.01
11	+ 先行 (先祖)	89.03
12	+ 先行 (子)	89.00
13	+ 先行 (子孫)	89.00
14	+ 先行 (all relation)	89.01
15	+ 全関係 (2,4,5,7,8,9,14)	<b>90.25</b>

た役割の数が少ないために比較的小さな向上に止まったものの、有意な上昇を示した。また、従来のラベル置き換えによる方法を役割名集合で試したところ、集合を素性関数として追加した場合よりも 0.5 ポイントほど精度が下がる事が示された。このことは、我々の集合素性を用いるアプローチが分類精度に貢献した事を示す。

#### 6.4 疎データ問題に対する効果

図 1 では、Micro F1 の値の変化は大きく見られないが、Macro F1 では役割集合の導入前と導入後に顕著な差が現れることから、これらの集合は役割分類における疎データ問題を効果的に解消している事が分かる。図 4 には実例数ごとに役割を分けて分類精度を示した。ここからも、我々の役割集合が特に実例の少ない役割に関して分類を助けている事が分かる。

#### 6.5 役割間関係のタイプ別効果

我々はまた、役割間の階層関係を用いた集合について、関係の型による効果の違いと先祖の影響を調べた。表 5 はそれぞれの Micro F1 を示したものである。特に継承と使用でその他の関係よりも効果が得られた。それ以外のものは、関係

の定義数そのものが少なかったために、それぞれ 0~0.05 ポイントの上昇をしたものの、効果の違いを考察するに至らなかった。また、親を持つ役割について、一代祖先の役割集合だけを発火させるよりも、祖先全ての集合を発火させた方がより効果がある事が確認できた。さらに、全ての型を区別せず混合して作った集合が個々の効果よりも比較的大きいことから、我々のアプローチにおいては関係の型を区別せずに取り扱い、より多くの情報共有を計る方がよいと言える。

## 7 まとめ

本稿では FrameNet の役割を適切にグループ化するための四つの方法を提案し、これらの集合を同時に扱うための素性関数を導入した。四種類の役割集合はそれぞれに分類の性能を向上させ、またこれらの異なる観点、異なる粒度の集合を重ねて加えることで、一つの観点による交わり合わない集合で役割ラベルを置き換える方法よりも分類精度が向上することを確認した。また、これらの集合が役割分類の疎データ問題を効果的に解消することを示した。この結果は役割が持つ幾つかの異なる性質を抽出したことによる。このことはよりの確な意味役割付与のための、役割の共通性を分解する仕組みの可能性と必要性を示唆する。また、我々の結果は PropBank においても同様のアプローチで精度の向上が得られる可能性を示唆する。

## 参考文献

- C. Baker, M. Ellsworth, and K. Erk. 2007. Semeval-2007 task 19: Frame semantic structure extraction. In *Proc. of SemEval'07*.
- U. Baldewein, K. Erk, S. Padó, and D. Prescher. 2004. Semantic role labeling with similarity based generalization using EM-based clustering. In *Proc. of Senseval'04*.
- C. A. Bejan, A. Moschitti, P. Morărescu, G. Nicolae, and S. Harabagiu. 2004. Semantic parsing based on framenet. In *Proc. of Senseval-3*.
- X. Carreras and L. Márquez. 2005. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proc. of CoNLL'05*.
- E. Charniak and M. Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proc. of ACL'05*.
- M. Ciaramita and Y. Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proc. of EMNLP'06*.
- R. Florian, S. Cucerzan, C. Shafer, and D. Yarowsky. 2003. Combining classifiers for word sense disambiguation. *Natural Language Engineering*, 8(04):327-341.
- D. Gildea and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245-288.
- A.-M. Giuglea and A. Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In *Proc. of COLING-ACL'06*.
- P. Kingsbury and M. Palmer. 2002. From Treebank to PropBank. In *Proc. of LREC'02*.
- A. Moschitti, A.-M. Giuglea, B. Coppolayz, and R. Basili. 2005. Hierarchical semantic role labeling. In *Proc. of CoNLL'05 shared task*.
- J. Nocedal. 1980. Updating quasi-newton matrices with limited storage. *Mathematics of Computation*, 35(151):773-782.
- S. Pradhan, K. Hacioglu, W. Ward, J. H. Martin, and D. Jurafsky. 2005. Semantic role chunking combining complementary syntactic views. In *Proc. of CoNLL'05 shared task*.
- J. Ruppenhofer, M. Ellsworth, M. Petruck, C. Johnson, and J. Scheffczyk. 2006. Framenet ii: Extended theory and practice. *Berkeley FrameNet*, Release 1.3.
- L. Shi and R. Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *Proc. of CILing'05*. Springer.
- M. Surdeanu, S. M. Harabagiu, J. Williams, and J. Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proc. of ACL'03*.
- N. Xue and M. Palmer. 2004. Calibrating features for semantic role labeling. In *Proc. of EMNLP'04*.
- B. Zupairain, E. Agirre, and L. Márquez. 2008. Robustness and generalization of role sets: Propbank vs. verbnet. In *Proc. of ACL'08*.