

# 自动浅层语义分析

车万翔, 刘挺, 李生

哈尔滨工业大学计算机学院信息检索实验室, 哈尔滨 150001

E-mail: {car, tliu, lisheng}@ir-lab.org

**摘要:** 全自动的语义分析一直是自然语言理解的主要目标之一。通过深层语义分析, 可以将自然语言转化为形式语言, 从而使计算机能够与人类无障碍的沟通。为达此目的, 人们已经进行了多年的努力, 然而目前取得的效果并不理想。浅层语义分析是对深层语义分析的一种简化, 它只标注与句子中谓词有关的成份的语义角色, 如施事、受事、时间和地点等。其能够对问答系统、信息抽取和机器翻译等应用产生推动作用。语义角色标注是浅层语义分析的一种实现方式, 具有定义清晰, 便于评价的优点。本文描述了浅层语义分析现有的语料库资源, 各种分析方法, 以及我们的工作等。并提出了对该任务一些难点问题的解决方案和对浅层语义分析发展的一个初步展望。

**关键词:** 自然语言理解; 浅层语义分析; 语义角色标注

## Shallow Semantic Parsing

Wanxiang Che, Ting Liu, Sheng Li

Information Retrieval Lab, School of Computer Science and Technology, Harbin Institute of Technology,

Harbin 150001

**Abstract:** Automatic semantic parsing is one of the main tasks for the natural language understanding. The natural language sentences can be translated into formal language by deep semantic parsing. Consequently computer and human beings can communicate with each other freely. In order to achieve the dream, people have done lots of efforts for many years. However the results are not up to much. Shallow semantic parsing is a simplified form of deep semantic parsing. It only labels the constituents with semantic roles which have direct relation with the predicate in a sentence. The semantic roles include Agent, Patient, Temporal, Locative and so on. In addition, it can give great support to many NLP applications, such as information extraction, question and answering, machine translation and so on. Semantic role labeling (SRL) is one kind of shallow semantic parsing. It is currently a well defined task with a substantial body of work and comparative evaluation. In the paper, we investigate the corpus for SRL, the parsing methods, and our previous work. At last, some proposals to solving the difficulties in shallow semantic parsing and some future work are given.

**Key words:** Natural language understanding; Shallow semantic parsing; Semantic role labeling

## 0 引言

所谓语义分析, 指的是将自然语言句子转化为反映这个句子意义(即句义)的某种形式化表示。即将人类能够理解的自然语言转化为计算机能够理解的形式语言, 做到人与机器的

互相沟通。

对句子进行正确的语义分析，一直是从事自然语言理解研究的学者们追求的主要目标。然而，限于目前的技术水平，深层的语义分析很难做到。人们目前更关心“浅层语义分析（Shallow Semantic Parsing）”，一种简化了的语义分析方式。它不考虑时态信息，例如“他将来北京。”与“他来北京了。”，虽然时态并不相同，但是浅层语义表示是相同的，同为：“来（他，北京）”。同时，浅层语义分析也不考虑谓词改变但语义不变的情况，例如“他出生于1969年3月18日。”与“他的生日是1969年3月18日。”，虽然它们的语义相同，但是浅层句法分析的结果并不一样，需要根据具体的应用进行更深入的处理。最后，浅层语义分析不考虑人、物的指代等情况。

语义角色标注（Semantic Role Labeling）是浅层语义分析的一种实现方式，具有分析任务定义明确，便于评价等优点。该方法并不对整个句子进行详细的语义分析，而只是标注句子中的一些成份为给定谓词（动词、名词等）的语义角色（参数），这些成分作为此谓词框架的一部分被赋予一定的语义含义，例如“[委员会 Agent][明天 Temp]将要[通过, \][此议案 Patient]。”其中，“通过”为谓词，“委员会”、“此议案”和“明天”分别是其施事、受事和发生的时间。这句话也可以表示成：

[明天 Temp][委员会 Agent]将要[通过, \][此议案 Patient]。

[明天 Temp][此议案 Patient]将要 被 [委员会 Agent][通过, \]。

[此议案 Patient][明天 Temp]将要 被 [委员会 Agent][通过, \]。

.....

这些句子虽然形式不同，但是它们的含义是一样的，也就是说谓词“通过”的各个参数相同。

进行语义角色标注的基础技术，如词性标注、句法分析、命名实体识别、统计机器学习等目前已经比较成熟。同时其在问答系统、信息抽取、机器翻译等领域有着广泛的应用<sup>1 2 3</sup>。以问答系统为例，语义角色标注指明某一活动发生的时间、地点等语义角色，自然的可以对这些类问题进行回答。因此目前语义角色标注引起了越来越多从事自然语言理解研究和应用的学者们的重视。

本文在接下来的部分首先介绍目前较为成熟的语义角色标注语料库以及这些语料库各自的特点；接着介绍了国际上相关的评测；然后，我们给出了目前进行语义角色标注的一些方法，以及我们的工作；最后，展望了今后语义角色标注技术的发展趋势。

## 1 语料资源

要想进行语义角色标注，需要好的语料资源的支持。目前，英语较为知名的浅层语义分析资源为 FrameNet<sup>4 5</sup>、PropBank<sup>6 7</sup>和 NomBank<sup>8 9</sup>。

其中，U.C.Berkeley 开发的 FrameNet 以框架语义<sup>10</sup>为标注的理论基础对英国国家语料库进行标注。它试图描述一个词汇单元（动词和部分名词以及形容词）的框架，同时也试图描述这些框架之间的关系。

PropBank 是 UPenn 在 Penn TreeBank 句法分析的基础上，标注的浅层语义信息。与

FrameNet 不同的是, PropBank 只对动词(非系动词)进行标注,相应的被称作目标动词。与 FrameNet 相比, PropBank 基于 Penn TreeBank 手工标注的句法分析结果,因此标注的结果几乎不受句法分析错误的影响,准确率较高;而且它几乎对 Penn TreeBank 中的每个动词及其语义角色进行了标注,因此覆盖范围更广,可学习性更强。

与 PropBank 标注 Penn TreeBank 中的动词做谓词不同, NomBank 标注了其中的名词作谓词的情况,参数的类别和表示同 PropBank 是一样的。例如:名词短语“John’ s replacement Ben”和“Ben’ s replacement of John”中,名词 replacement 便是谓词, Ben 是 ARG0,表示替代者, John 是 ARG1 表示被替代者。

除英语外,许多其它语言也建立了各自的语义角色标注库,例如: SALSA<sup>11</sup>是德语版的 FrameNet; Prague Dependency Treebank<sup>12</sup>项目进行了大量的句法和语义标注(捷克语),甚至包括指代消解的标注等; Chinese PropBank<sup>13</sup>以及 Chinese NomBank<sup>14</sup>分别是 Upenn 基于 Chinese Penn TreeBank<sup>15</sup>标注的,和 PropBank 以及 NomBank 对应的,并且动词和名词作谓词的汉语语义角色标注资源。

## 2 语义角色标注评测

### 2.1 评价方法

在语义角色标注中,通常采用信息检索中使用的 F-Score 来评价系统的性能,定义为:

$$F-Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

其中 Precision 是准确率, Recall 是召回率,定义分别为:

$$Precision = \frac{\text{正确标注为语义角色的个数}}{\text{分类器预测为语义角色总数}}, \quad Recall = \frac{\text{正确标注为语义角色的个数}}{\text{测试数据中语义角色总数}}$$

### 2.2 国际评测

对于语义角色标注,国际上曾经举行过三次评测,分别为 Senseval-3<sup>16</sup>以及 CoNLL(Conference on Computational Linguistics Learning)会议主办的 SRL(Semantic Role Labeling) Share Task 2004<sup>17</sup>以及 2005<sup>18</sup>。其中 Senseval-3 是以 FrameNet 为训练和测试语料;而 CoNLL Shared Task 则是以 PropBank 为语料库。2004 年来自 Colorado University 的 Hacioglu 等人<sup>19</sup>,采取以短语为标注单元,语义角色识别和分类分步进行的策略,使用 SVM 算法在不使用全局特征的条件下,获得了最好的标注结果,测试集合的 F-Score 达到了 69.49%;2005 年来自 UIUC 的 Koomen 等人<sup>20</sup>使用 SNoW 分类器,综合多种深层句法分析的输出结果,加上使用整数线性规划(Integer Linear Programming)的后处理方法,取得了最好的成绩,测试集合的 F 值达到了 79.44%。这也代表了当今最好的语义角色标注效果。

3 自动语义角色标注方法

有了语料库资源作为基础，接下来的，也是最主要的任务就是从这些资源中学习已有知识，然后进行自动的浅层语义分析。

3.1 标注基本单元

自动标注的基本单元可以是句法成分 (Constituent)、短语 (Phrase)、词 (Word) 或者依存关系 (Dependency Relation) 等等。

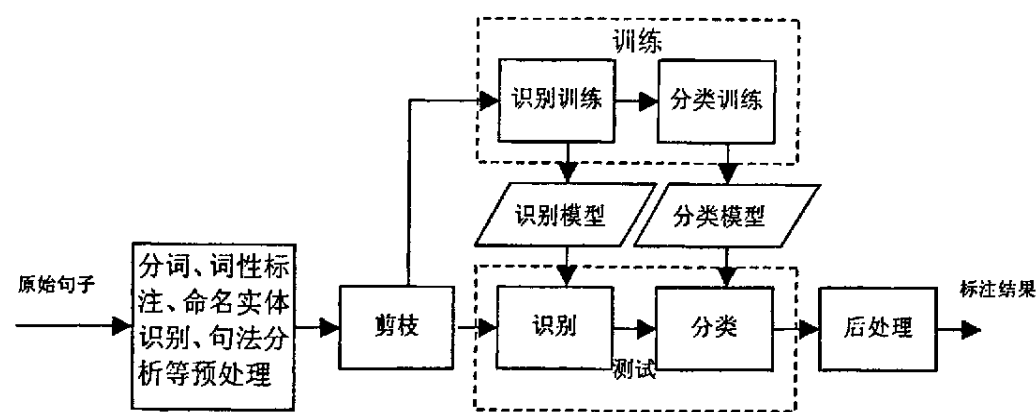
一般认为每个语义角色是与某一句法成分相对应的。也就是说一个语义角色必然对应着一个句法成分，反之未必。因此，现在多数的语义角色标注系统通常都是以句法成分为基本的标注单元的。这种策略，在句法分析比较成熟的语言（如英文等）上表现得较好。

然而，在其它语言上，很难自动的获得这种深层句法分析的结果，而且现有的句法分析系统，在通用领域表现欠佳。为此有人试图将浅层语义分析建立在浅层句法分析的基础之上，毕竟浅层句法分析的鲁棒性要好于深层句法分析。因为通过浅层句法分析只能获得非嵌套短语的信息，而不能获得全部的句法分析结果，也就不能获得句法成分的分析结果，但是我们一般认为一个非嵌套的短语属于同一语义角色，因此产生了使用短语作为浅层语义分析基本单元的系统<sup>21</sup>。词是比短语更细的语言单位，有些浅层语义分析系统也使用其作为标注的基本单位，然而并不如基于短语和句法成分的效果理想<sup>22</sup>。

除此以外，Hacioglu 使用依存句法分析结果进行浅层语义分析<sup>23</sup>，也取得了与基于短语结构句法分析方法相近的结果。

3.2 标注过程

通常的语义角色标注分为 4 个步骤：剪枝 (Pruning)、识别 (Identification)、分类 (Classification) 和后处理 (Post-processing)。如图 1 所示：



图（1）语义角色标注过程

其中，剪枝是指在基于短语结构的语义角色标注系统中，根据启发式规则，删除大部分

不可能成为语义角色的句法成分<sup>24</sup>，这样可以大幅减少输入到分类器中的实例个数，提高训练和测试的效率。识别过程一般是对一个标注单元是否是语义角色加以判别，并保留识别成语义角色的标注单元，待进一步分类究竟属于哪个语义角色类，这样也可以减少进入分类判别的实例的个数，加快处理速度。最后根据语义角色标注本身的一些固有约束进行后处理<sup>25</sup>。这些约束包括，一个谓语动词不能有重复的语义角色并且语义角色不存在嵌套等等。Sang等人又提出了一种类似拼写检查方法的后处理方法<sup>26</sup>，就是将最终标注的结果与谓语动词框架文件进行比较，去除不可能的语义标注角色。

### 3.3 基于特征的学习方法

我们可以将识别和分类步骤看成是一个基于复杂条件的分类问题。最初解决此类问题是由人类专家根据自身的经验总结规则知识，构建知识库，再使用这些知识进行分类。然而，这种方法不但需要专业领域专家的知识和花费专家们大量的时间，而且规则之间会产生冲突。

为了克服知识库方法的缺点，人们后来使用统计学习的方法。该方法不需要有专业技能的专家书写知识库，只需要有一定专业知识的人对数据按照学习目标进行一定的标注；然后以此为训练数据，构造统计学习模型；再使用各种统计学习方法对模型参数加以求解。

在语义角色的识别和分类过程中，一般是根据一定的语言学知识列出标注单元的各种特征，并与该单元的语义角色类型（可能不属于任何语义角色）组成学习的实例，最后使用某种学习算法对这些实例进行自动的学习，以对新的实例进行预测。

语义角色标注中经常使用的统计学习方法包括最大熵<sup>27</sup>、支持向量机(SVM)<sup>27</sup>、感知器等<sup>28</sup>。一般的，这些方法又被称作基于特征的方法，因为在使用它们的时候，需要人为的定义大量具有区分度的特征模板，然后根据此模板将每个实例转化为特征向量进行学习或预测操作。

### 3.4 特征选取

在语义角色标注中，有一些基本的语言学特征对系统的性能有着直接的影响。这些特征及其说明举例如下：

**谓词：**即待标注的目标动词，通常取其原形；

**路径：**指的是待标注的单元与目标动词之间句法分析树上的路径。

**短语类型：**待标注的单元所属短语的类型，如 NP、VP、PP 等；

**位置：**待标注的单元与目标动词之间的相对位置，这是二元特征，分别为“前”和“后”；

**语态：**目标动词的语态二元特征，分别为“主动”和“被动”；

**核心词：**标注单元的核心词；

**动词子类框架：**记录了目标动词及其子节点的结构信息。

以上这些特征，都从不同的侧面反映了待标注单元的语义角色信息。但是这些仅仅反映了待标注单元的局部信息。一些全局特征，如标注单元语义角色的序列特征，全句含有的目标动词数目等，也可以作为特征加入。

### 3.5 基于 Kernel 的学习方法

基于 Kernel 的机器学习方法,是与常用的基于特征的机器学习方法相对应的一种方法。其基本思想是将低维线性不可分问题映射到高维空间,变为线性可分问题。通常这种映射可以通过计算 Kernel 函数隐式的达到,从而降低时间和空间复杂性。Kernel 函数的计算能很好的融入支持向量机、感知器等学习算法,使它们可以较好的处理线性不可分问题。

在自然语言处理领域,研究者们又将 Kernel 的思想升华,使其能够更好的处理结构化的数据,发掘更多的结构化特征。最早由 Haussler<sup>29</sup>以及 Watkins<sup>30</sup>提出使用动态规划的方法高效计算结构化数据的 Kernel 函数,这又被称作 Convolution Kernel。在此之后,越来越多的学者利用相似的思想于自己的研究工作,如句法分析<sup>31</sup>,文本分类<sup>32</sup>等等。特别是 Moschitti 通过计算句法分析树的 Kernel 函数进行语义角色标注<sup>33</sup>。

## 4 我们的工作

### 4.1 基于最大熵的语义角色标注

在语义角色标注过程中,并非所有的系统都包括剪枝、识别、分类和后处理 4 个步骤,特别是剪枝步骤的主要目的是提高处理效率,但随之带来的是召回率的下降,损失了一些本应是语义角色的标注单元。同样的情况也发生在识别步骤中。因此,在 CoNLL 2005 语义角色标注评测中,我们去除了剪枝步骤,并且合并了识别和分类步骤,直接对语义角色进行分类,采用最大熵分类器,使用若干基本特征和扩展特征及其组合<sup>34</sup>。获得了并列的第六名的成绩,并且是采用单一句法分析系统中的第二名。最终在性能如表 1 所示。

表 (1) CoNLL 2005 Shared Task 实验结果

	Precision	Recall	F-Score
Development	79.65%	71.34%	75.49
Test WSJ	81.30%	73.37%	77.13
Test Brown	71.65%	60.36%	65.52
Test WSJ+Brown	80.02%	71.65%	75.60

### 4.2 混合卷积树 Kernel

在基于 Kernel 的方法中,经过我们的研究发现, Moschitti 提出的 PAF (Predicate Argument Feature) Kernel 并不能很好的刻画语义角色标注的特征。因此,我们提出了一种新的 Kernel 计算方法 - 混合卷积树 Kernel (Hybrid Convolution Tree Kernels)<sup>35</sup>,如图 2 所示。该 Kernel 充分考虑了句法结构自身的特征以及句法结构与谓词之间关系的特征,并且通过将这两种特征进行合理的融合。

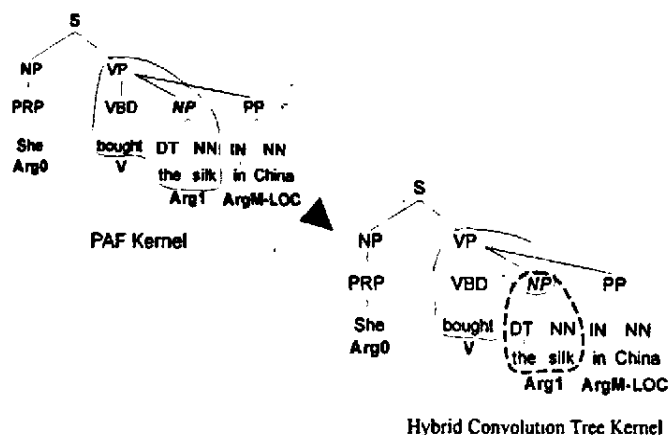


图 (2): PAF Kernel与Hybrid Convolution Tree Kernel的比较

该方法与PAF Kernel在开发集上F值的对比结果如表2所示:

表 (2) 两种Kernel方法对比

	Hybrid	PAF
Development Set	66.01%	64.38%

#### 4.3 中文语义角色标注

目前, 中文语义角色标注主要以 Chinese PropBank 以及 Chinese NomBank 为分析对象<sup>37</sup>, 借鉴英文分析的方法, 也是基于短语结构句法分析的结果。我们使用 Chinese PropBank 的手工句法分析数据, 采用最大熵分类器, 加入额外特征及组合, F 值达到了 91.31%。与使用基础特征的性能 (F=89.76%) 相比, 提高了约 1.3 个百分点。

然而, 我们认为短语结构句法分析的结果, 对于表达语义角色信息, 并不直观。例如, 对于句子“他在北京开出租车。”, 有如图 3 所示的短语结构句法分析结果以及对应的语义角色标注结果:

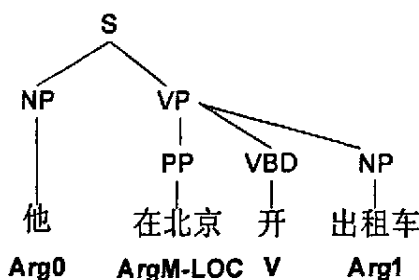
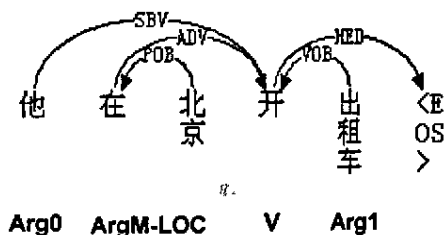


图 (3): 短语结构句法分析结果以及对用的语义角色标注结果

直观上, “他”以及“出租车”应该是与谓词“开”有类似的句法结构关系, 但是短语结构句法分析的结果并非如此。相反, 基于依存句法分析的结果恰好能弥补这一不足, 同样例子, 依存句法分析结果如图 4 所示:



图（4）：依存句法分析结果以及对用的语义角色标注结果

可见，依存句法分析中，核心词动词或名词一般是语义角色的谓语，并且每个核心词与其支配词的关系含有关系类型，这种类型对于分析语义角色很有帮助。

基于以上的考虑，我们认为利用我实验室开发的“语言技术平台（LTP）”<sup>1</sup> 很适合作为汉语语义角色标注的基础模块。LTP 是一个中文处理的集成平台，将断句、分词、词性标注、命名实体识别、词义消歧、句法分析等自然语言处理的基础模块通过 XML 语言进行了很好融合。

## 5 研究展望

语义角色标注是对自然语言理解的一个有益的探索，它以其它自然语言处理技术为基础，使用统计学习的方法进行分析。因为解决此问题需要丰富的知识，所以需要对问题做更加详细的分析，提取更为有效的特征，使用更为适合的模型。

### 5.1 克服数据稀疏问题

目前，语义角色标注的研究才刚刚起步，可以利用的人工加工资源还很有限，尤其是汉语等语言，可供学习的语料库非常有限。因此，与其它自然语言处理问题相比，语义角色标注遇到的数据稀疏问题更加严重。

通过分析现有系统以及对浅层语义分析任务的深入认识，我们认为要想解决数据稀疏问题，对词汇（包括谓词以及参数）进行适当的聚类是必然出路。例如，“我爱吃苹果。”与“我爱吃香蕉。”中“苹果”和“香蕉”都是“吃”的受事，也就是它们的语义角色相同，如果能将类似的词语进行聚类，就可以克服一定的数据稀疏问题，提高系统的泛化能力。基于此，我们可以使用 LTP 平台中的词义消歧（Word Sense Disambiguation, WSD）技术，同时我们也可以试图寻找泛化能力更强的特征，以提高系统的召回率，同时提高系统的性能。

### 5.2 语义角色标注应用

如前所述，语义角色标注对如问答系统等多种自然语言处理应用技术有着很强的支撑作用。根据我实验室的实际研究需要，我们以信息抽取中的事件抽取为应用背景，阐述中文语

<sup>1</sup> <http://ltp.ir-lab.org/>



义角色标注的应用方法。

根据国际 ACE (Automatic Content Extraction) 会议的定义, 事件抽取需要找到相应的触发动词, 发生的时间、地点, 参与的人物等信息, 一般一个事件范围不超过一个句子。可见这种事件抽取方式, 利用语义角色标注可以很好的完成, 我们需要找出每个事件对应的触发动词, 然后标注出它的各种语义角色作为事件的参数, 并根据参数的特点, 进一步确认是否为需要的事件。这里需要注意的是, 在实际应用中, 参数往往是较长的人名、地名、机构名、时间和数字等, 因此, 对这些实体的识别变得异常重要, 同时若实体的识别准确率提高, 语义角色标注的性能也将随之提高。

如本文开始所述, 人们对于自然语言理解的最终目标是真正的深层次的语义分析, 以期进行自动的知识获取, 推理等等。因此在浅层语义分析的基础上, 进行深层的语义分析将成为未来研究的重点。为达此目标, 我们必须建立性能更加卓越的浅层语义分析系统。

## 参考文献

- <sup>1</sup> S. Narayanan and S. Harbabagiu. Question answering based on semantic structures. In *Proc. Coling 2004*, 2004.
- <sup>2</sup> M. Surdeanu, S. Harabagiu, J. Williams, and P. Aarseth. Using predicate-argument structures for information extraction. In *Proc. ACL 2003*, 2003.
- <sup>3</sup> N. Xue and S. Kulick. Automatic predicate argument structure analysis of the penn chinese treebank. In *Proc. Machine Translation Summit IX*, 2003.
- <sup>4</sup> <http://framenet.icsi.berkeley.edu/>
- <sup>5</sup> C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet project. In C. Boitet and P. Whitelock, editors, *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pages 86-90, San Francisco, California, 1998. Morgan Kaufmann Publishers.
- <sup>6</sup> <http://www.cis.upenn.edu/~ace/>
- <sup>7</sup> M. Palmer, D. Gildea, and P. Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, 31(1), 2005.
- <sup>8</sup> <http://nlp.cs.nyu.edu/meyers/NomBank.html>
- <sup>9</sup> A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. Annotating noun argument structure for NomBank. In *Proceedings of LREC-2004*, pages 803-806, Lisbon, Portugal.
- <sup>10</sup> C. J. Fillmore. Frames and the semantics of understanding. *Quaderni di Semantica*, IV(2), 1985.
- <sup>11</sup> K. Erk, A. Kowalski, S. Pado, and M. Pinkal. Towards a resource for lexical semantics: A large german corpus with extensive semantic annotation. In *Proceedings of at ACL 2003*, Sapporo, 2003.
- <sup>12</sup> E. Hajicova. Prague dependency treebank: From analytic to tectogrammatical annotation. In *Proceedings of the First Workshop on Text, Speech, Dialogue*, pages 45-50, 1998.
- <sup>13</sup> <http://www.cis.upenn.edu/~chinese/cpb/>
- <sup>14</sup> N. Xue. 2006. Annotating the predicate-argument structure of Chinese nominalizations. In *Proceedings of the*

- <sup>15</sup> <http://www.cis.upenn.edu/~chinese/>
- <sup>16</sup> <http://www.cs.unt.edu/~rada/senseval/senseval3/workshop.html>
- <sup>17</sup> X. Carreras and L. Màrques. Introduction to the conll-2004 shared task: Semantic role labeling. In *Proceedings of CoNLL-2004*, pages 89-97, Boston, MA, USA, 2004.
- <sup>18</sup> <http://www.lsi.upc.edu/~srlconll/>
- <sup>19</sup> K. Hacioglu, S. Pradhan, W. Ward, J. H. Martin, and D. Jurafsky. Semantic role labeling by tagging syntactic chunks. In *Proceedings of CoNLL-04*, 2004.
- <sup>20</sup> P. Koomen, V. Punyakanok, D. Roth, et al. Generalized Inference with Multiple Semantic Role Labeling Systems. *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, Michigan: Association for Computational Linguistics, 2005, 181-184.
- <sup>21</sup> K. Hacioglu, S. Pradhan, W. Ward, et al. Semantic Role Labeling by Tagging Syntactic Chunks. H. T. Ng, E. Riloff, (Editors) *HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, USA: Association for Computational Linguistics, 2004, 110-113
- <sup>22</sup> O. Y. Kwong, B. K. Tsou. Semantic Role Tagging for Chinese at the Lexical Level. *Proceedings of IJCNLP 2005*, 2005
- <sup>23</sup> K. Hacioglu. Semantic Role Labeling Using Dependency Trees. *Proc. Coling 2004*, 2004
- <sup>24</sup> N. Xue, M. Palmer. Calibrating Features for Semantic Role Labeling. *Proc. EMNLP 2004*, 2004
- <sup>25</sup> V. Punyakanok, D. Roth, W. Yih, et al. Semantic Role Labeling via Integer Linear Programming Inference. *COLING-2004*, Geneva, Switzerland, 2004, 1346-1352.
- <sup>26</sup> E. Tjong Kim Sang, S. Canisius, A. van den Bosch, et al. Applying Spelling Error Correction Techniques for Improving Semantic Role Labelling. *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, Michigan: Association for Computational Linguistics, 2005, 229-232.
- <sup>27</sup> S. Pradhan, K. Hacioglu, V. Krugler, et al. Support Vector Learning for Semantic Argument Classification. *Machine Learning Journal*. 2005
- <sup>28</sup> X. Carreras, L. Marquez, G. Chrupala. Hierarchical Recognition of Propositional Arguments with Perceptrons. H. T. Ng, E. Riloff, (Editors) *HLTNAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, USA: Association for Computational Linguistics, 2004, 106-109
- <sup>29</sup> D. Haussler. Convolution Kernels on Discrete Structures. Tech. Rep. UCSC-CRL-99-10, 1999.
- <sup>30</sup> C. Watkins. Dynamic Alignment Kernels. Tech. Rep. CSD-TR-98-11, 1999.
- <sup>31</sup> M. Collins, N. Duffy. Convolution Kernels for Natural Language. *Proceedings of NIPS-2001*, 2001.
- <sup>32</sup> H. Lodhi, C. Saunders, J. Shawe-Taylor, et al. Text classification using string Kernels. *J. Mach. Learn. Res.* 2002, 2:419-444
- <sup>33</sup> A. Moschitti. A Study on Convolution Kernels for Shallow Semantic Parsing. *Proc. ACL 2004*, 2004
- <sup>34</sup> T. Liu, W. Che, S. Li, et al. Semantic Role Labeling System Using Maximum Entropy Classifier. *Proceedings*

---

of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), Ann Arbor, Michigan: Association for Computational Linguistics, 2005, 189–192.

<sup>35</sup> Wanxiang Che, Min Zhang, Ting Liu, Sheng Li. A Hybrid Convolution Tree Kernel for Semantic Role Labeling. In *Proceedings of ACL2006*. Sydney. July 2006

<sup>36</sup> 刘怀君, 车万翔, 刘挺. 中文语义角色标注的特征工程, SWCL-2006. 2006.

<sup>37</sup> Nianwen Xue. Semantic Role Labeling of nominalized predicates in Chinese, in *Proceedings of HTL-NAACL 2006*. New York City. 2006.

<sup>38</sup> 郎君, 刘挺, 张会鹏, 李生, LTP:语言技术平台, SWCL-2006, 2006.

作者：[车万翔](#)，[刘挺](#)，[李生](#)  
作者单位：[哈尔滨工业大学计算机学院信息检索实验室, 哈尔滨, 150001](#)

本文读者也读过(10条)

1. [陈耀东](#), [王挺](#), [陈火旺](#) [浅层语义分析研究](#)[会议论文]-2007
2. [邹娟](#), [周经野](#), [邓成](#), [Zou Juan](#), [Zhou Jingye](#), [Deng Cheng](#) [一种基于语义分析的中文特征值提取方法](#)[期刊论文]-[计算机工程与应用](#)2005, 41 (36)
3. [车万翔](#), [刘挺](#), [李生](#) [浅层语义分析](#)[会议论文]-
4. [朱译翔](#) [中文名词性谓词的语义角色标注的研究](#)[期刊论文]-[福建电脑](#)2010, 26 (9)
5. [俞士汶](#), [王治敏](#), [朱学锋](#) [文学语言与自然语言理解研究](#)[会议论文]-2006
6. [周巧云](#), [ZHOU Qiao-yun](#) [面向计算机的深度语义分析](#)[期刊论文]-[喀什师范学院学报](#)2009, 30 (2)
7. [张亮](#), [王树梅](#), [黄河燕](#), [张孝飞](#), [ZHANG Liang](#), [WANG Shu-mei](#), [HUANG He-yan](#), [ZHANG Xiao-fei](#) [面向中文问答系统的问句句法分析](#)[期刊论文]-[山东大学学报（理学版）](#) 2006, 41 (3)
8. [丁金涛](#) [基于特征向量的语义角色标注研究](#)[学位论文]2008
9. [张巍](#), [陈俊杰](#), [ZHANG Wei](#), [CHEN Junjie](#) [浅层语义分析及SPARQL在问答系统中的应用](#)[期刊论文]-[计算机工程与应用](#)2011, 47 (2)
10. [张亮](#), [王树梅](#), [黄河燕](#), [张孝飞](#) [面向中文问答系统的问句句法分析](#)[会议论文]-2006

引用本文格式：[车万翔](#), [刘挺](#), [李生](#) [自动浅层语义分析](#)[会议论文] 2006