

# 词汇概念结构与空间语义表示

王希豪

北京大学中国语言文学系

*wangxihao@pku.edu.cn*

2021 年 12 月 2 日

- 1 国际空间语义理解工作坊 (SpLU)
- 2 词汇概念结构 (Lexical Conceptual Structures, LCS)
- 3 The Case for Systematically Derived Spatial Language Usage[Dorr and Voss, 2018a]

# International Workshop on Spatial Language Understanding (SpLU)

## 会议介绍

- 2018 年 NAACL 期间第一次举办
- 2019 年和 2021 年与 Workshop on Language Grounding for Robotics (RoboNLP) 联合举办
- 主要关注
  - Spatial Language Meaning Representation (Continuous, Symbolic)
  - Spatial Language Learning
  - Spatial Language Reasoning
  - Combining Vision and Language for Spatial Understanding
  - .....

# International Workshop on Spatial Language Understanding (SpLU)

## 组织者

- Parisa Kordjamshidi
- James Pustejovsky
- Archana Bhatia
- Marie-Francine Moens
- .....

# International Workshop on Spatial Language Understanding (SpLU)

## 组织者

- Parisa Kordjamshidi
  - 密歇根州立大学助理教授
  - 佛罗里达人类与机器认知研究所 (IHMC) 研究员
  - 2013 年于天主教鲁汶大学获得了博士学位，博士论文为 *Structured Machine Learning for Mapping Natural Language to Spatial Ontologies*，博士导师是 Marie-Francine Moens
  - SpRL、mSpRL、SpartQA 任务的提出者和构建者之一
- James Pustejovsky
- Archana Bhatia
- Marie-Francine Moens
- .....

# International Workshop on Spatial Language Understanding (SpLU)

## 组织者

- Parisa Kordjamshidi
- James Pustejovsky
  - 布兰迪斯大学教授
  - 生成词库理论的提出者
  - TimeML, ISO-Space 的提出者
  - Semantic Annotation Framework (semAF) 的提出者
  - *Handbook of Linguistic Annotation* 的主编
- Archana Bhatia
- Marie-Francine Moens
- .....

# International Workshop on Spatial Language Understanding (SpLU)

## 组织者

- Parisa Kordjamshidi
- James Pustejovsky
- Archana Bhatia
- Marie-Francine Moens
- .....

Parisa Kordjamshidi, James Pustejovsky, Marie-Francine Moens 三人在 2020 年的 EMNLP 上报告了题为 *Representation, Learning and Reasoning on Spatial Language for Downstream NLP Tasks* 的 tutorial

[https://virtual.2020.emnlp.org/tutorial\\_T5.html](https://virtual.2020.emnlp.org/tutorial_T5.html)

<https://spatial-language-tutorial.github.io/>

- ① 国际空间语义理解工作坊 (SpLU)
- ② 词汇概念结构 (Lexical Conceptual Structures, LCS)
- ③ The Case for Systematically Derived Spatial Language Usage[Dorr and Voss, 2018a]

Jackendoff [1983, 1990, 1996] 提出的 LCS 是一种组合式的抽象语义表示框架。相比于 AMR, EDS, UCCA 等语义表示框架, LCS 对动词的论元结构进行了较为细致的描写, 还刻画了动词的空间语义信息。

Jackendoff [1983, 1990, 1996] 提出的 LCS 是一种组合式的抽象语义表示框架。相比于 AMR, EDS, UCCA 等语义表示框架, LCS 对动词的论元结构进行了较为细致的描写, 还刻画了动词的空间语义信息。

基于 LCS, 研究者们开发了一系列的应用:

- Dorr [1993], Habash and Dorr [2002] 以 LCS 作为中间语言设计了基于规则和统计的机器翻译系统
- Levow et al. [2000] 使用 LCS 进行跨语言的信息抽取
- Traum and Habash [2001] 把 LCS 应用到了文本生成任务中
- Dorr [1997] 利用 LCS 构造词典和进行二语习得指导
- .....

## 基本结构 [Traum and Habash, 2001]

LCS 是带根的有向图，图中的每个结点都包含特定的信息

LCS 是带根的有向图，图中的每个结点都包含特定的信息<sup>1</sup>

- 类型 (Type)
  - 事件 (Event)
  - 状态 (State)
  - 路径 (Path)
  - 方式 (Manner)
  - 特性 (Property)
  - 物体 (Thing)

---

<sup>1</sup>这里各个概念的中文翻译可能不是很严谨

# 基本结构 [Traum and Habash, 2001]

LCS 是带根的有向图，图中的每个结点都包含特定的信息<sup>1</sup>

- 类型 (Type)
- 基础谓词 (Primitive predicate)
  - 闭合类 (closed class primitive): CAUSE, GO, BE, TO, ...
  - 开放类 (open class primitive):  
reduce+ed, textile+, slash+ingly, ...

---

<sup>1</sup>这里各个概念的中文翻译可能不是很严谨

# 基本结构 [Traum and Habash, 2001]

LCS 是带根的有向图，图中的每个结点都包含特定的信息<sup>1</sup>

- 类型 (Type)
- 基础谓词 (Primitive predicate)
- 领域 (Field)
  - 表示空间与运动的 (Loc)ational
  - 表示领属的 (Poss)etional
  - 表示时间的 (Temp)oral
  - 表示状态的 (Ident)ificational
  - 表示情景的 (Circ)umstantial
  - 表示存在的 (Exist)ential
  - 表示感知的 (Perc)eptual
  - 表示交流的 (Comm)unicational
  - 表示工具的 (Instr)umental
  - 表示意图的 (Intent)ional

---

<sup>1</sup>这里各个概念的中文翻译可能不是很严谨

# 基本结构 [Traum and Habash, 2001]

LCS 是带根的有向图，图中的每个结点都包含特定的信息<sup>1</sup>

- 类型 (Type)
- 基础谓词 (Primitive predicate)
- 领域 (Field)

LCS 的组合性

- 句子中的每个词都有对应的 Root LCS (RLCS)
- 短语的 LCS 是由多个 RLCS 组合得到的 Composed LCS (CLCS)

---

<sup>1</sup>这里各个概念的中文翻译可能不是很严谨

RLCS: 削減

```
(act_on loc (* thing 1) (* thing 2)
  ((* [on] 23) loc (*head*) (thing 24))
  (cut+ingly 26)
  (down+/m))
```

## 例子 [Traum and Habash, 2001]

CLCS: 美国单方面削减了中国纺织品的出口额度

```
(cause (us+)
  (go ident (quota+ (china+)
    (textile+)
    (export+))
  (to ident (quota+ (china+)
    (textile+)
    (export+))
  (at ident (quota+ (china+)
    (textile+)
    (export+))
  (reduce+ed))))
(with instr (*HEAD*) nil)
(unilaterally+/m))
```

表示: The United States caused the quota to go identificationally towards being at the state of being reduced

## 原理 - 理论基础 [Levin and Hovav, 2011]

20 世纪 80 年代，一些语言学家提出了“句子语法结构主要的部分是从词语的词汇特征映射得到的”的思想。

在一些研究工作中，研究者会假设“动词伴有谓词-论元结构”，从而可以得到推论：“动词的谓词-论元结构投射到句子中，形成了特定的句法结构”。

## 原理 - 理论基础 [Levin and Hovav, 2011]

20 世纪 80 年代，一些语言学家提出了“句子语法结构主要的部分是从词语的词汇特征映射得到的”的思想。

在一些研究工作中，研究者会假设“动词伴有谓词-论元结构”，从而可以得到推论：“动词的谓词-论元结构投射到句子中，形成了特定的句法结构”。

### 存在的问题

- 动词的论元数量是可变的
  - a Pat dried the clothes.
  - b The clothes dried.
- 同样论元数的动词可以表现出不同的句法形式
  - c Pat sold the rare book to Terry.
  - d Pat sold Terry the rare book.

### 存在的问题

- 动词的论元数量是可变的
  - a Pat dried the clothes.
  - b The clothes dried.
- 同样论元数的动词可以表现出不同的句法形式
  - c Pat sold the rare book to Terry.
  - d Pat sold Terry the rare book.

### LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围

## 存在的问题

- 动词的论元数量是可变的
  - a Pat dried the clothes.
  - b The clothes dried.
- 同样论元数的动词可以表现出不同的句法形式
  - c Pat sold the rare book to Terry.
  - d Pat sold Terry the rare book.

## LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围
- 简单来说，就是将动词按照意义（模板）分类，不同的类有不同的论元/句法结构

### LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围
- 简单来说，就是将动词按照意义（模板）分类，不同的类有不同的论元/句法结构

### 例子 ([come to be STATE])

- a break: y come to be BROKEN
- b break: x cause (y come to be BROKEN)

## LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围
- 简单来说，就是将动词按照意义（模板）分类，不同的类有不同的论元/句法结构

## 例子 ([come to be STATE])

- a break: y come to be BROKEN
- b break: x cause (y come to be BROKEN)

## 例子

- a I cut the rope around his wrists.  
cut: x produce CUT in y, by sharp edge coming into contact with y
- b I cut at the rope around his wrists.  
cut: x causes sharp edge to move along path toward y, in order to produce CUT on y, by sharp edge coming into contact with y

### LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围
- 简单来说，就是将动词按照意义（模板）分类，不同的类有不同的论元/句法结构

### 补充

- LCS 不关注动词的具体意义如何解释，它关注的是重复出现的、可以决定动词句法表现的那部分意义。因此，*freeze* 和 *melt* 都属于状态改变类动词，具有同样的 LCS。

## LCS 的基础假设

- 在动词的论元结构之上，我们可以分离出一小部分重复出现的意义成分，这些成分决定了动词论元数量的范围
- 简单来说，就是将动词按照意义（模板）分类，不同的类有不同的论元/句法结构

## 补充

- LCS 不关注动词的具体意义如何解释，它关注的是重复出现的、可以决定动词句法表现的那部分意义。因此，*freeze* 和 *melt* 都属于状态改变类动词，具有同样的 LCS。
- 在具体的实现中，LCS 的意义表示采用了谓词分解 (predicate decomposition) 的思想，将复杂的意义表示为基础谓词的组合。例子中，*cut* 的意义解释可分解为“move”和“produce”两个部分，其中，“move”可以用基础谓词 GO 来解释，“produce”则可表示为基础谓词 CAUSE 和 BECOME 的组合。

## 原理 - 基础谓词 [Levin and Hovav, 2011]

基础谓词是 LCS 体系中，表达意义的最基本单位<sup>2</sup>

- 不可以通过某种方式转化成为其他基础谓词
- 更不可以由其他基础谓词定义

---

<sup>2</sup>类比向量空间中的基

基础谓词是 LCS 体系中，表达意义的最基本单位<sup>2</sup>

- 不可以通过某种方式转化成为其他基础谓词
- 更不可以由其他基础谓词定义

基本假设

- 我们可以限定基础事件 (类型) 的数量，使得其他事件都可以被分析成为这些基础事件
- 每个基础事件包含有限个基础谓词，每个基础谓词表示了这类事件的一个基本类型

---

<sup>2</sup>类比向量空间中的基

基础谓词是 LCS 体系中，表达意义的最基本单位<sup>2</sup>

- 不可以通过某种方式转化成为其他基础谓词
- 更不可以由其他基础谓词定义

基本假设

- 我们可以限定基础事件 (类型) 的数量，使得其他事件都可以被分析成为这些基础事件
- 每个基础事件包含有限个基础谓词，每个基础谓词表示了这类事件的一个基本类型

Jackendoff [1983, 1990] 接受了 localist hypothesis, 认为:

- 运动 (motion) 事件和位置 (location) 事件是最基础的事件类型，其他事件都可以由这两类事件组合得到

---

<sup>2</sup>类比向量空间中的基

Jackendoff [1983, 1990] 接受了 localist hypothesis, 认为:

- 运动 (motion) 事件和位置 (location) 事件是最基础的事件类型, 其他事件都可以由这两类事件组合得到
  - 运动事件可以用基础谓词 GO 表示  
GO 含有 Theme 和 Path 这两个论元
  - 方位事件可以用基础谓词 BE(表状态) 和 STAY(表非状态) 表示  
BE 和 STAY 都含有论元 Theme 和 Location
  - 除此之外, 基础谓词 CAUSE 和 LET 用来组合表示复杂的事件

Jackendoff [1983, 1990] 接受了 localist hypothesis, 认为:

- 运动 (motion) 事件和位置 (location) 事件是最基础的事件类型, 其他事件都可以由这两类事件组合得到
- 有些事件并不能直接显式地转化成为运动或是位置事件, 但可以表示成为抽象的运动或位置事件

## 原理 - 基础谓词 [Levin and Hovav, 2011]

Jackendoff [1983, 1990] 接受了 localist hypothesis, 认为:

- 运动 (motion) 事件和位置 (location) 事件是最基础的事件类型, 其他事件都可以由这两类事件组合得到
- 有些事件并不能直接显式地转化成为运动或是位置事件, 但可以表示成为抽象的运动或位置事件
  - 领属事件: 所有物可以看作是 Theme, 所有者可以看作是 Location
  - 物理/精神状态的变化:  
“位于”一种状态的实体“移动”到了另一种状态

例子 ([CAUSE (x, (STAY y, z))])

- a Terry kept the bike in the shed.
- b Terry kept the bike basket.
- c Terry kept the bike clean.

例子 ([CAUSE (x, (STAY y, z))])

- a Terry kept the bike in the shed.
- b Terry kept the bike basket.
- c Terry kept the bike clean.

问题

- 在这里，keep 都表示“保持”的意思，句子也具有相似的结构，但实际上，它所描述的事件却是很不相同的。我们怎么才能区分它所描述的事件呢？

例子 ([CAUSE (x, (STAY y, z))])

- a Terry kept the bike in the shed. (Locational)
- b Terry kept the bike basket. (Possessional)
- c Terry kept the bike clean. (Identificational)

问题

- 在这里，keep 都表示“保持”的意思，句子也具有相似的结构，但实际上，它所描述的事件却是很不相同的。我们怎么才能区分它所描述的事件呢？ ⇒ **领域 (Field)**

Levin [1993] 根据动词的意义和句法表现，将 3024 个动词的 4186 个义项分成了 192 个类。

Levin [1993] 根据动词的意义和句法表现，将 3024 个动词的 4186 个义项分成了 192 个类。

Dorr et al. [2001] 在 Levin [1993] 的基础上，添加了 44 个新的类，并将原有的 192 个类细化，将 9525 个动词义项划分成了 500 个类。同时，还标注出了每个动词的 LCS 结构。

## 资源 - LCS 动词数据库 (LVD)[Dorr et al., 2001]

Levin [1993] 根据动词的意义和句法表现, 将 3024 个动词的 4186 个义项分成了 192 个类。

Dorr et al. [2001] 在 Levin [1993] 的基础上, 添加了 44 个新的类, 并将原有的 192 个类细化, 将 9525 个动词义项划分成了 500 个类。同时, 还标注出了每个动词的 LCS 结构。

### 例子

词项: REDUCE

词类编号: 45.4.a

```
LCS: (cause (* thing 1)
      (go ident (* thing 2)
        (toward ident (thing 2)
          (at ident (thing 2)
            (reduce+ed 9))))
      ((* with 19) instr (*head*) (thing 20)))
```

Levin [1993] 根据动词的意义和句法表现，将 3024 个动词的 4186 个义项分成了 192 个类。

Dorr et al. [2001] 在 Levin [1993] 的基础上，添加了 44 个新的类，并将原有的 192 个类细化，将 9525 个动词义项划分成了 500 个类。同时，还标注出了每个动词的 LCS 结构。

### 例子

词项: PUT

词类编号: 9.1

```
LCS: (cause (* thing 1)
      (go loc (* thing 2)
        ((* to 5) loc (thing 2)
          ([at] loc (thing 2) (thing 6))))))
```

## 资源 - LCS 动词数据库 (LVD)[Dorr et al., 2001]

Levin [1993] 根据动词的意义和句法表现，将 3024 个动词的 4186 个义项分成了 192 个类。

Dorr et al. [2001] 在 Levin [1993] 的基础上，添加了 44 个新的类，并将原有的 192 个类细化，将 9525 个动词义项划分成了 500 个类。同时，还标注出了每个动词的 LCS 结构。

### 例子

```
LVD Class Entry:
(:NUMBER "51.1.a"
 :NAME "Verbs Inherently Directed Motion / -from/to"
 :WORDS (advance arrive ascend climb come depart descend enter escape
         exit fall flee go leave plunge recede return rise tumble)
 :NON_LEVIN_WORDS (approach come! defect head)
 :LCS (go loc (* thing 2)
       ((* from 3) loc (thing 2) (at loc (thing 2) (thing 4)))
       ((* to 5) loc (thing 2) ([at] loc (thing 2) (thing 6)))
       (!!+ingly 26))
 :THETA_ROLES ((1 "_th,src(from),goal(towards)")
               (1 "_th,goal,src(from)")
               (1 "_th,src(from),goal(to)"))))
```

图: LVD 条目: 51.1.a

- ① 国际空间语义理解工作坊 (SpLU)
- ② 词汇概念结构 (Lexical Conceptual Structures, LCS)
- ③ The Case for Systematically Derived Spatial Language Usage[Dorr and Voss, 2018a]

## Bonnie Dorr

- 马里兰大学计算机科学系及高级计算机研究所教授
- 佛罗里达人类与机器认知研究所 (IHMC) 高级研究员
- 1990 年于麻省理工学院获得博士学位，博士论文为 *Lexical Conceptual Structure and Machine Translation*，导师 Robert C. Berwick 是 Noam Chomsky 和 Marvin Minsky 的学生
- 1993 年在博士论文的基础上，出版了 *Machine Translation: A View from the Lexicon*，提出了一种基于 GB 理论和 LCS 理论的语际机器翻译 (interlingual machine translation) 系统
- 在基于规则的机器翻译系统逐渐没落之后，主要关注数据驱动的机器翻译系统研究

## 研究背景

- 此前，空间语言理解的相关任务（机器人导航等）关注于如何将自然语言映射到深度概念化的、非语言学的表示，但是，在机器人导航、人类-机器人交流等任务当中，又往往缺少足够的资源进行大规模的计算。

## 研究背景

- 此前，空间语言理解的相关任务（机器人导航等）关注于如何将自然语言映射到深度概念化的、非语言学的表示，但是，在机器人导航、人类-机器人交流等任务当中，又往往缺少足够的资源进行大规模的计算。

⇒ 如何将现有的词汇-语义资源利用起来，去解决空间语言理解中的上述问题？

## 研究背景

- 此前，空间语言理解的相关任务（机器人导航等）关注于如何将自然语言映射到深度概念化的、非语言学的表示，但是，在机器人导航、人类-机器人交流等任务当中，又往往缺少足够的资源进行大规模的计算。

⇒ 如何将现有的词汇-语义资源利用起来，去解决空间语言理解中的上述问题？

⇒ 系统化地从已有资源中提取空间理解可用的模式

## 研究背景

- 此前，空间语言理解的相关任务（机器人导航等）关注于如何将自然语言映射到深度概念化的、非语言学的表示，但是，在机器人导航、人类-机器人交流等任务当中，又往往缺少足够的资源进行大规模的计算。

⇒ 如何将现有的词汇-语义资源利用起来，去解决空间语言理解中的上述问题？

⇒ 系统化地从已有资源中提取空间理解可用的模式

## LCS 的优点

- 具有组合性的词汇形式语义表示框架
- 作为中间语言被应用到了机器翻译和信息抽取任务当中，有在其他语言上应用的潜力

## 句法-语义接口

- 动词和名词性谓词的内在词汇-语义结构与它们的论元结构有着密切的关系 [Jackendoff, 1996, Dorr, 1993, Levin, 1993, Kipper et al., 2008, Palmer et al., 2017]
- 介词和动词词类之间的联系可以作为语义内容预测的重要因素 [Kipper et al., 2004]
- 英语中介词的生成既依赖于谓词的语义，又依赖于谓词在句法-语义接口上的结构特性 [Voss et al., 1995]

## 句法-语义接口

- 动词和名词性谓词的内在词汇-语义结构与它们的论元结构有着密切的关系 [Jackendoff, 1996, Dorr, 1993, Levin, 1993, Kipper et al., 2008, Palmer et al., 2017]
- 介词和动词词类之间的联系可以作为语义内容预测的重要因素 [Kipper et al., 2004]
- 英语中介词的生成既依赖于谓词的语义，又依赖于谓词在句法-语义接口上的结构特性 [Voss et al., 1995]

⇒ 谓词-论元结构句法-语义接口的重要组成部分，除此之外，介词也是联系句法和语义的桥梁。

## 句法-语义接口

- 动词和名词性谓词的内在词汇-语义结构与它们的论元结构有着密切的关系 [Jackendoff, 1996, Dorr, 1993, Levin, 1993, Kipper et al., 2008, Palmer et al., 2017]
- 介词和动词词类之间的联系可以作为语义内容预测的重要因素 [Kipper et al., 2004]
- 英语中介词的生成既依赖于谓词的语义，又依赖于谓词在句法-语义接口上的结构特性 [Voss et al., 1995]

⇒ 谓词-论元结构句法-语义接口的重要组成部分，除此之外，介词也是联系句法和语义的桥梁。

## 句法-语义映射的基础

- 意义成分 (Component of meaning) - 隐式的意义成分 (语义层面)，例如，动词 *elevate* 就具有 UPWARD 的意义
- 组合成分 (Collocation) - 特定词的显式补充成分 (句法层面)，比如 “lift up” 中的 *up*

## 句法-语义映射的基础

- 意义成分 (Component of meaning) - 隐式的意义成分 (语义层面), 例如, 动词 *elevate* 就具有 UPWARD 的意义
- 组合成分 (Collocation) - 特定词的显式补充成分 (句法层面), 比如 “lift up” 中的 *up*

## 语言使用模式

- 阻碍 (Block)
- 重合 (Overlap)
- 填充 (Fill)

## 句法-语义映射的基础

- 意义成分 (Component of meaning) - 隐式的意义成分 (语义层面), 例如, 动词 *elevate* 就具有 UPWARD 的意义
- 组合成分 (Collocation) - 特定词的显式补充成分 (句法层面), 比如 “lift up” 中的 *up*

## 语言使用模式

- 阻碍 (Block): 在语义上包含方位相关意义成分的词语, 不和对应的组合成分共同出现。例如, *elevate* 和和意义成分相对应的配置不会出现的词语, 比如 *elevate* 和 *ascend*, 在语义上都表示向上 (UPWARD), 但在句子实现时不需要介词来表现方位意义
- 重合 (Overlap)
- 填充 (Fill)

## 句法-语义映射的基础

- 意义成分 (Component of meaning) - 隐式的意义成分 (语义层面), 例如, 动词 *elevate* 就具有 UPWARD 的意义
- 组合成分 (Collocation) - 特定词的显式补充成分 (句法层面), 比如 “lift up” 中的 *up*

## 语言使用模式

- 阻碍 (Block)
- 重合 (Overlap): 在语义上包含方位相关意义成分的词语, 在表现到句法上时, 可以和对应的组合成分共现 (但不强制)。比如, *lift* 和 *raise*, 在语义上同样有向上 (UPWARD) 的意义, 在使用时依然可以和 *up* 组成词组
- 填充 (Fill)

## 句法-语义映射的基础

- 意义成分 (Component of meaning) - 隐式的意义成分 (语义层面), 例如, 动词 *elevate* 就具有 UPWARD 的意义
- 组合成分 (Collocation) - 特定词的显式补充成分 (句法层面), 比如 “lift up” 中的 *up*

## 语言使用模式

- 阻碍 (Block)
- 重合 (Overlap)
- 填充 (Fill): 在语义上不包含方位相关意义成分的词语
  - a Oblig(atory) - 如果没有组合成分就无法确定动作的方向, 且组合成分一定要出现, 比如, Put the box onto the table
  - b Opt(ional) - 如果没有组合成分就无法确定动作的方向, 但组合成分可以不出现, 比如, Move the box (to the table)

## 语言使用模式

- 阻碍 (Block)
- 重合 (Overlap)
- 填充 (Fill)

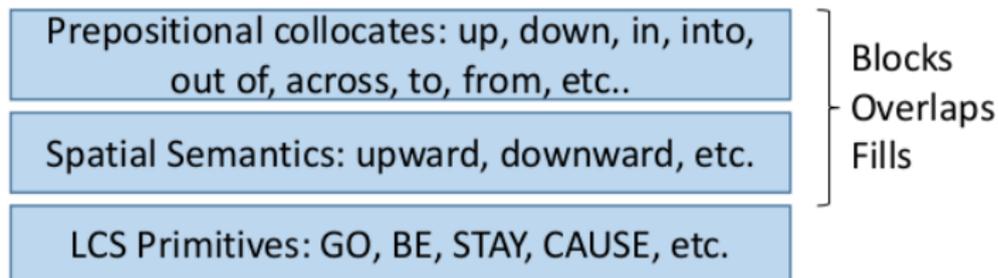


图: 基于 LCS 的空间语义理解、生成的三个层面

# 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

## LVD Class Entry:

```
(:NUMBER "51.1.a"  
:NAME "Verbs Inherently Directed Motion / -from/to"  
:WORDS (advance arrive ascend climb come depart descend enter escape  
        exit fall flee go leave plunge recede return rise tumble)  
:NON_LEVIN_WORDS (approach come! defect head)  
:LCS (go loc (* thing 2)  
      ((* from 3) loc (thing 2) (at loc (thing 2) (thing 4)))  
      ((* to 5) loc (thing 2) ([at] loc (thing 2) (thing 6)))  
      (!!+ingly 26))  
:THETA_ROLES ((1 "_th,src(from),goal(towards)")  
              (1 "_th,goal,src(from)")  
              (1 "_th,src(from),goal(to)")))
```

图: STYLUS 条目: 51.1.a

# 资源 - SysTematicallY Derived Language USage (STYLUS)[Dorr and Voss, 2018b]

## STYLUS Class Entry:

```
(:NUMBER "51.1.a"  
:NAME "Verbs Inherently Directed Motion / -from/to"  
:WORDS (advance arrive ascend climb come depart descend enter escape  
        exit fall flee go leave plunge recede return rise tumble)  
:NON_LEVIN_WORDS (approach come! defect head)  
:COLLOCATIONS ("from" "to" "towards")  
:COMPONENT_OF_MEANING (FROM, AWAY_FROM, OUT_OF, OR, UP_TO, UP,  
        BEFORE, INTO, TO)) ;; Fill-Opt <<SPATIAL>>
```

图: STYLUS 条目: 51.1.a

# 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

## STYLUS Class Entry:

```
(:NUMBER "51.1.a"  
:NAME "Verbs Inherently Directed Motion / -from/to"  
:WORDS (advance arrive ascend climb come depart descend enter escape  
        exit fall flee go leave plunge recede return rise tumble)  
:NON_LEVIN_WORDS (approach come! defect head)  
:COLLOCATIONS ("from" "to" "towards")  
:COMPONENT_OF_MEANING (FROM, AWAY_FROM, OUT_OF, OR, UP_TO, UP,  
        BEFORE, INTO, TO)) ;; Fill-Opt <<SPATIAL>>
```

图: STYLUS 条目: 51.1.a

STYLUS 继承了简化版的 LVD 词类, 省略了完整的 LCS 结构和  $\theta$ -Role

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”
- 确定用法类型

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”
- 确定用法类型
  - 如果“意义成分”不为空但“组合成分”为空，属于 Block  
例如: Face the doorway

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”
- 确定用法类型
  - 如果“意义成分”不为空但“组合成分”为空，属于 Block  
例如: Face the doorway
  - 如果“意义成分”和“组合成分”都不为空，属于 Overlap  
例如: Follow (behind) the car

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - Systematically Derived Language Usage (STYLUS) [Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”
- 确定用法类型
  - 如果“意义成分”不为空但“组合成分”为空，属于 Block  
例如: Face the doorway
  - 如果“意义成分”和“组合成分”都不为空，属于 Overlap  
例如: Follow (behind) the car
  - 如果“意义成分”和“组合成分”都不为空，且  $\theta$ -Role 中指出介词必须出现，属于 Fill-Oblig  
例如: Put the box onto the table

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

# 资源 - SysTematicallY Derived Language USage (STYLUS)[Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类, 省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取 “意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取 “组合成分”
- 确定用法类型
  - 如果 “意义成分” 不为空但 “组合成分” 为空, 属于 Block  
例如: Face the doorway
  - 如果 “意义成分” 和 “组合成分” 都不为空, 属于 Overlap  
例如: Follow (behind) the car
  - 如果 “意义成分” 和 “组合成分” 都不为空, 且  $\theta$ -Role 中指出介词必须出现, 属于 Fill-Oblig  
例如: Put the box onto the table
  - 如果 “意义成分” 为空, “组合成分” 为空或者  $\theta$ -Role 中指出介词不一定要出现, 则属于 Fill-Opt  
例如: Move the box (to the table)

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

## 资源 - SysTematicallY Derived Language USage (STYLUS)[Dorr and Voss, 2018b]

STYLUS 继承了简化版的 LVD 词类，省略了完整的 LCS 结构和  $\theta$ -Role 自动构造过程

- 从 CatVat 数据库 [Habash and Dorr, 2003] 中抽取“意义成分”<sup>2</sup>
- 从  $\theta$ -Role 中抽取“组合成分”
- 确定用法类型
- 确定是否为空间类型
  - 如果 LCS 结构属于有 Locational 领域，
  - 或者 LCS 结构属于 Possetional 领域，但伴有 GO, BE 等基础论元，那么该动词被标记为空间类型；
  - 否则该动词就不是空间类型的

---

<sup>2</sup>这个数据库是在 LVD 基础上构建的

# 资源 - SysTematicallY Derived Language USage (STYLUS)[Dorr and Voss, 2018b]

Lexical Ops	LVD Classes	Spatial Subset	Spatial Examples	Non-Spatial Subset	Non-Spatial Examples
Block	7	7	elevate, face, pocket	0	infect, archive
Overlap	17	10	advance, lower, lift	7	follow, precede
Fill-Oblig	310	128	drive, rotate, put	182	mount, install
Fill-Opt	87	49	remove, slide	38	move, add
Intrans	6	3	float, part, squirm	3	choke, drown, snap
N/A	73	12	—	61	—
<b>Tot. Classes</b>	500	219		281	
<b>Tot. Verbs</b>	9525	4640		4885	

图: STYLUS 数据统计

# 资源 - SysTematicallY Derived Language USage (STYLUS)[Dorr and Voss, 2018b]

Class [Pattern]	Robot Navigation		Cyber Notification (Generation Only)
	Understanding	Generation	
9.8-Fill [Block]	<i>Face</i> the doorway <i>Face</i> to the doorway	Where do I <i>face</i> ? *Where do I <i>face</i> to?	Risk that trojan horse virus will <i>infect</i> (*to) System A increased 5%
47.8.1-Contig [Overlap]	<i>Follow</i> the car Follow behind the car	Which car do I <i>follow</i> (behind)?	Risk increased 25% that VPN failure will <i>follow</i> (behind) malware attack
9.1-Put [FillOblig]	<i>Put</i> it on the table * <i>Put</i> the box	Where do I <i>put</i> it? *Do I <i>put</i> the box?	Risk of malware increases 5% if disk A <i>mounted</i> on file server B
11.2-Slide [FillOpt]	<i>Move</i> it to the table <i>Move</i> it	Where do I <i>move</i> it? Do I <i>move</i> it?	10% increased malware risk if System A files are <i>moved</i> (to System B)

图：用法模板展示

# 总结

- LCS 不仅表示了动词的论元结构，还对动词的空间语义进行了刻画

# 总结

- LCS 不仅表示了动词的论元结构，还对动词的空间语义进行了刻画
- LCS 作为中间表示，已经被应用到了跨语言的文本生成任务当中。尽管基于 LCS 的机器翻译系统很难覆盖许多语言现象，但如果只是应用到空间方位语句生成这一小的范围内，或许能有不错的表现。特别是在机器人导航等领域，可以考虑利用 LCS 生成简单语句。

# 总结

- LCS 不仅表示了动词的论元结构，还对动词的空间语义进行了刻画
- LCS 作为中间表示，已经被应用到了跨语言的文本生成任务当中。尽管基于 LCS 的机器翻译系统很难覆盖许多语言现象，但如果只是应用到空间方位语句生成这一小的范围内，或许能有不错的表现。特别是在机器人导航等领域，可以考虑利用 LCS 生成简单语句。
- Bonnie J. Dorr 基于 LCS 构建了一些语义数据库，但主要还是对词汇的刻画，如果想要在此基础上进行研究，目前可能只能考虑基于规则的方法，应该不太能直接应用基于统计和神经网络的技术。

# 总结

- LCS 不仅表示了动词的论元结构，还对动词的空间语义进行了刻画
- LCS 作为中间表示，已经被应用到了跨语言的文本生成任务当中。尽管基于 LCS 的机器翻译系统很难覆盖许多语言现象，但如果只是应用到空间方位语句生成这一小的范围内，或许能有不错的表现。特别是在机器人导航等领域，可以考虑利用 LCS 生成简单语句。
- Bonnie J. Dorr 基于 LCS 构建了一些语义数据库，但主要还是对词汇的刻画，如果想要在此基础上进行研究，目前可能只能考虑基于规则的方法，应该不太能直接应用基于统计和神经网络的技术。
- Dorr and Voss [2018a] 从 LCS 中根据介词和空间语义提取出了三类语言使用模板，对于简单空间方位语句的生成具有指导意义。作者在论文中还提到，可以考虑研究如何自动地从 AMR, PDT 等数据中自动获取语言使用模板。然而，作者的方法更多地是基于 LCS 的特性，我认为难以扩展到其他语义表示框架中。

- B. Dorr and C. Voss. The case for systematically derived spatial language usage. In Proceedings of the First International Workshop on Spatial Language Understanding, pages 63–70, New Orleans, June 2018a. Association for Computational Linguistics. doi: 10.18653/v1/W18-1408. URL <https://aclanthology.org/W18-1408>.
- B. Dorr and C. Voss. Stylus: A resource for systematically derived language usage. In Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing, pages 57–64, 2018b.
- B. J. Dorr. Machine translation: a view from the lexicon. MIT press, 1993.
- B. J. Dorr. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. Machine Translation, 12 (4):271–322, 1997.

## 参考文献 II

- B. J. Dorr, M. Olsen, N. Habash, and S. Thomas. Lcs verb database documentation., 2001. URL [http://users.umiacs.umd.edu/~bonnie/Demos/LCS\\_Database\\_Documentation.html](http://users.umiacs.umd.edu/~bonnie/Demos/LCS_Database_Documentation.html).
- N. Habash and B. Dorr. Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation. In Conference of the Association for Machine Translation in the Americas, pages 84–93. Springer, 2002.
- N. Habash and B. Dorr. A categorial variation database for english. Technical report, MARYLAND UNIV COLLEGE PARK INST FOR ADVANCED COMPUTER STUDIES, 2003.
- R. Jackendoff. Semantics and cognition. 1983.
- R. Jackendoff. Semantic structures. 1990.
- R. Jackendoff. The proper treatment of measuring out, telicity, and perhaps even quantification in english. Natural Language & Linguistic Theory, 14(2):305–354, 1996.

- K. Kipper, B. Snyder, and M. Palmer. Using prepositions to extend a verb lexicon. In Proceedings of the Computational Lexical Semantics Workshop at HLT-NAACL 2004, pages 23–29, 2004.
- K. Kipper, A. Korhonen, N. Ryant, and M. Palmer. A large-scale classification of english verbs. Language Resources and Evaluation, 42 (1):21–40, 2008.
- B. Levin. English verb classes and alternations: A preliminary investigation. University of Chicago press, 1993.
- B. Levin and M. R. Hovav. 19. lexical conceptual structure. In Volume 1, pages 420–440. De Gruyter Mouton, 2011.
- G.-A. Levow, B. J. Dorr, and D. Lin. Construction of chinese-english semantic hierarchy for information retrieval. Technical report, MARYLAND UNIV COLLEGE PARK INST FOR ADVANCED COMPUTER STUDIES, 2000.

- M. Palmer, C. Bonial, and J. D. Hwang. Verbnet: Capturing english verb behavior, meaning and usage. The Oxford handbook of cognitive science, pages 315–336, 2017.
- D. Traum and N. Habash. Generation from lexical conceptual structures. Technical report, MARYLAND UNIV COLLEGE PARK INST FOR ADVANCED COMPUTER STUDIES, 2001.
- C. R. Voss, B. J. Dorr, and M. U. Sencan. Lexical allocation in interlingua-based machine translation of spatial expressions. In InWorking Notes for IJCAI-95 Workshop on the Representation and Processing of Spatial Expressions, Montreal, Canada. Citeseer, 1995.