The Internal Structure of Coordinate Categories

Roger Levy

UC San Diego
Department of Linguistics
Presented at Structure and Evidence in Linguistics

29 April 2013
The historical arc of constraint-based grammar

- Hallmark principles of the Sag tradition of grammatical analysis:
The historical arc of constraint-based grammar

- Hallmark principles of the Sag tradition of grammatical analysis:
  - It’s essential to state formal claims about the nature of grammar precisely
The historical arc of constraint-based grammar

- Hallmark principles of the Sag tradition of grammatical analysis:
  - It’s essential to state formal claims about the nature of grammar precisely
  - We must seriously engage with the full range of distributional generalizations in the data
The historical arc of constraint-based grammar

- Hallmark principles of the Sag tradition of grammatical analysis:
  - It’s essential to state formal claims about the nature of grammar precisely
  - We must seriously engage with the full range of distributional generalizations in the data
  - But we must also be rigorous in determining when to attribute a distributional generalization to the grammar proper
The historical arc of constraint-based grammar

- Hallmark principles of the Sag tradition of grammatical analysis:
  - It’s essential to state formal claims about the nature of grammar precisely
  - We must seriously engage with the full range of distributional generalizations in the data
  - But we must also be rigorous in determining when to attribute a distributional generalization to the grammar proper

- When implemented correctly, these principles are powerful in identifying both grammatical knowledge and its interface with the rest of cognition
The internal structure of coordinate categories

- Principle of *Conjoin Likes* (Chomsky, 1965)
  \[ X \rightarrow X \text{ Conj } X \]
The internal structure of coordinate categories

- Principle of *Conjoin Likes* (Chomsky, 1965)
  \[ X \rightarrow X \text{ Conj } X \]

- Empirically false for gross syntactic category (Sag et al., 1985):

  *Pat is a Republican and proud of it (NP and AdjP)*
The internal structure of coordinate categories

- Principle of *Conjoin Likes* (Chomsky, 1965)
  \[ X \rightarrow X \text{ Conj } X \]

- Empirically false for gross syntactic category (Sag et al., 1985):
  
  *Pat is a Republican and proud of it (NP and AdjP)*

- And for case-marking (Przepiórkowski, 1999; Levy, 2001):
  
  *proždal* “waited” governs ACC or GEN

  Včera vec’ den’ on proždal \[ \text{[NP svoju yesterday all day he expected. ACC OR GEN self’s. ACC} \]
  podrugu Irinu i \[ \text{[NP zvonka ot svoego brata girlfriend. ACC Irina. ACC and call. GEN from self’s brother} \]
  Grigorija]. (Russian, Levy, 2001)

  Gregory

  “Yesterday he waited all day for his girlfriend Irina and for a call from his brother Gregory.”
What’s left for grammar?

- Generalization: a coordination is CATEGORICALLY GRAMMATICAL iff it satisfies all the extrinsic constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)
What’s left for grammar?

- Generalization: a coordination is **categorically grammatical** iff it satisfies all the *extrinsic* constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)

- So... was “Conjoin Likes” just wrong?
What’s left for grammar?

- Generalization: a coordination is **CATEGORICALLY GRAMMATICAL** iff it satisfies all the *extrinsic* constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)

- So... was “Conjoin Likes” just wrong?

- Is there anything left for grammar to say about a “tendency” for coordinated categories to be like one another?
Corpus data

- Unlike-category coordinations are easy to find in corpora

  His son had been friendly, a big fellow of fifty or more, a fishing-boat captain and powerful like the sea

  (Parsed Brown corpus)
Corpus data

- Unlike-category coordinations are easy to find in corpora
  
  *His son had been friendly, a big fellow of fifty or more, a fishing-boat captain and powerful like the sea*

- But there is a huge *quantitative* tendency for coordination to be of like categories in corpora

  *Right-hand conjunct*  
  
<table>
<thead>
<tr>
<th>NP</th>
<th>AdjP</th>
<th>1308</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-hand Conjunct</td>
<td>NP</td>
<td>6</td>
<td>114</td>
</tr>
</tbody>
</table>

  *(Parsed Brown corpus)*
Conjoin Likes as a gradient grammatical constraint?

- It is tempting to claim immediately that this pattern illustrates a “soft constraint” (one of Miller’s “Usage Preferences”) toward Conjoin Likes.
Conjoin Likes as a gradient grammatical constraint?

- It is tempting to claim immediately that this pattern illustrates a “soft constraint” (one of Miller’s “Usage Preferences”) toward Conjoin Likes.
- But should we really attribute this to the grammar proper?
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:

- Categorical: the *possibility* of a string is sufficient to demand the grammar account for it, regardless of the extralinguistic circumstances required
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:

- Probabilistic/gradient theories: the data currency is relative prevalence, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances.
**Conjoin Likes** as a gradient grammatical constraint?

Critical difference between nature of *evidence* for categorical versus probabilistic/gradient grammatical theories:

- Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances.
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:

- Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances.

```
S
  NP  VP
    V_{trans}  X
      Y and Z
  S
    NP  VP
      be  X
        Y and Z
  NP
    X  N
      Y and Z
```
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:

- Probabilistic/gradient theories: the data currency is relative prevalence, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances.

```
S
  NP  VP
    V_{trans}  X
      Y and Z
    ↓  NP and NP
```

```
S
  NP  VP
    be  X
      Y and Z
    ↓  Uncorrelated Mixture
```

```
NP
  X  N
    Y and Z
    ↓  AdjP and AdjP
```
Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of evidence for categorical versus probabilistic/gradient grammatical theories:

- Probabilistic/gradient theories: the data currency is relative prevalence, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances.
Directed Acyclic Graphical Models ("Bayes Nets")

Bayes Nets specify:

- Probabilistic conditional independencies: $X$ and $Y$ are conditionally independent given known variables iff every path between $X$ and $Y$ is blocked by:
  - an unknown variable with "converging arrows"; or
  - a known variable without "converging arrows"

The basic units of probabilistic (gradient) knowledge, $P_{p\text{child}|parents}$:

$P_{\text{Alarm}|\text{Earthquake}, \text{Burglary}}$

$P_{\text{Call}|\text{Alarm}}$

(Example due to Russell and Norvig, 2003)
Directed Acyclic Graphical Models ("Bayes Nets")

Bayes Nets specify:

- Probabilistic conditional independencies: $X$ and $Y$ are **CONDITIONALLY INDEPENDENT** given known variables iff every path between $X$ and $Y$ is blocked by:
  - an unknown variable **with** "converging arrows"; or

(Example due to Russell and Norvig, 2003)
Bayes Nets specify:

- Probabilistic conditional independencies: \( X \) and \( Y \) are CONDITIONALLY INDEPENDENT given known variables iff every path between \( X \) and \( Y \) is blocked by:
  - an unknown variable with "converging arrows"; or

(Example due to Russell and Norvig, 2003)
Directed Acyclic Graphical Models ("Bayes Nets")

Bayes Nets specify:

- Probabilistic conditional independencies: $X$ and $Y$ are CONDITIONALLY INDEPENDENT given known variables iff every path between $X$ and $Y$ is blocked by:
  - an unknown variable with “converging arrows”; or
  - a known variable without “converging arrows”

(Example due to Russell and Norvig, 2003)
Directed Acyclic Graphical Models ("Bayes Nets")

Bayes Nets specify:

- Probabilistic conditional independencies: $X$ and $Y$ are **CONDITIONALLY INDEPENDENT** given known variables iff every path between $X$ and $Y$ is blocked by:
  - an unknown variable **with** "converging arrows"; or
  - a known variable **without** "converging arrows"

(Example due to Russell and Norvig, 2003)
Directed Acyclic Graphical Models ("Bayes Nets")

Bayes Nets specify:

- Probabilistic conditional independencies: X and Y are CONDITIONALLY INDEPENDENT given known variables iff every path between X and Y is blocked by:
  - an unknown variable with "converging arrows"; or
  - a known variable without "converging arrows"

- The basic units of probabilistic (=gradient) knowledge, \( P(\text{child}|\text{parents}) \):
  
  \[
P(\text{Alarm}|\text{Earthquake, Burglary}) \\
P(\text{Call}|\text{Alarm})
\]

(Example due to Russell and Norvig, 2003)
Conjoin Likes in a probabilistic grammar

\[ M_1, M_2 \quad \text{Intended conjunct meanings and extrinsic constraints} \]
\[ F_1, F_2 \quad \text{Realized linguistic forms of the conjuncts} \]
\[ O \quad \text{Ordering decision} \]

(NB: Connections from \( M_i \) to \( O \) are necessary to account for semantic interpretive constraints pertaining to order, e.g., \( eat \text{ and run} \neq \text{run and eat} \); Cooper and Ross, 1975)
Conjoin Likes in a probabilistic grammar

\[ M_1, M_2 \] Intended conjunct meanings and extrinsic constraints
\[ F_1, F_2 \] Realized linguistic forms of the conjuncts
\[ O \] Ordering decision

(NB: Connections from \( M_i \) to \( O \) are necessary to account for semantic interpretive constraints pertaining to order, e.g., \textit{eat and run} \neq \textit{run and eat}; Cooper and Ross, 1975)
What “gradient coordination of like categories” means:

\[ P(F_1, F_2|M_1, M_2) \]

is especially high when \( F_1 \) and \( F_2 \) are “like” in the traditional sense of

\[
X \rightarrow X \text{ and } X
\]

Fully technically:

\[
pMI(F_1, F_2|M_1, M_2) = \log \frac{P(F_1, F_2|M_1, M_2)}{P(F_1|M_1)P(F_2|M_2)}
\]

is monotonically increasing in the structural similarity of \( F_1 \) and \( F_2 \).
Empirical prediction

If forms are gradiently “more grammatical” to the native speaker when they are more probable...
Empirical prediction

If forms are gradiently “more grammatical” to the native speaker when they are more probable.

...then like-category coordinations should be judged to be more natural, or acceptable, than unlike-category coordinations.
Experiment 1

Acceptability judgment study (scale of 1–9):

Pat is a Republican and a freak. \([\text{Noun Noun}]\)
Pat is a Republican and freaky. \([\text{Noun Adj}]\)
Pat is Republican and a freak. \([\text{Adj Noun}]\)
Pat is Republican and freaky. \([\text{Adj Adj}]\)

(Baseline: The children decorated the sparkling ornaments onto the tree was a 4.)
Experiment 1: Results

The gradient preference for coordination of unlike categories is pretty strong!
Greater explanatory power of gradient constraints

- We saw that “Conjoin Likes” is categorically false, but “probabilistically” true
Greater explanatory power of gradient constraints

- We saw that “Conjoin Likes” is categorically false, but “probabilistically” true
- But why stop at major syntactic categories—what about category-internal structure (Johnson, 1998; Klein and Manning, 2003)?
Greater explanatory power of gradient constraints

- We saw that “Conjoin Likes” is categorically false, but “probabilistically” true
- But why stop at major syntactic categories—what about category-internal structure (Johnson, 1998; Klein and Manning, 2003)?
- Such a grammatical preference has previously been explored under the rubric of parallelism (Frazier et al., 1984; Hale et al., 2006; Dubey et al., 2008)

```
NP
  NP
    ...α...
  Conj
and
NP
  ...α...
```
NP-internal parallelism: the genitive alternation

Postnominal
The future of our country ~ Our country’s future
The base of the lamp ~ The lamp’s base
The tail of a cat ~ A cat’s tail

Prenominal

The Internal Structure of Coordinate Categories
Corpus data on genitive alternation parallelism

<table>
<thead>
<tr>
<th>Left Conjunct</th>
<th>Right Conjunct</th>
<th>Post</th>
<th>Pre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>77</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>20</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

- There is also strong evidence for a parallelism preference in the genitive alternation...
Corpus data on genitive alternation parallelism

<table>
<thead>
<tr>
<th>Left Conjunct</th>
<th>Right Conjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post</td>
</tr>
<tr>
<td>Post</td>
<td>77</td>
</tr>
<tr>
<td>Pre</td>
<td>20</td>
</tr>
</tbody>
</table>

- There is also strong evidence for a parallelism preference in the genitive alternation...
- ...but once again this analysis fails to control for conjunct meanings $M_1, M_2$
Corpus data on genitive alternation parallelism

<table>
<thead>
<tr>
<th>Left Conjunct</th>
<th>Right Conjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post</td>
</tr>
<tr>
<td>Post</td>
<td>77</td>
</tr>
<tr>
<td>Pre</td>
<td>20</td>
</tr>
</tbody>
</table>

- There is also strong evidence for a parallelism preference in the genitive alternation.
- . . .but once again this analysis fails to control for conjunct meanings $M_1, M_2$
- We can control this more tightly with an experiment
Experiment 2

Acceptability judgment study (scale of 1–9):

Terry assembled...

...the frame of the chair and the base of the lamp. \[Post \ Post\]

...the frame of the chair and the lamp’s base. \[Post \ Pre\]

...the chair’s frame and the base of the lamp. \[Pre \ Post\]

...the chair’s frame and the lamp’s base. \[Pre \ Pre\]

(Baseline: The children decorated the sparkling ornaments onto the tree was a 4.)
There is also a preference for parallelism among realizations of the genitive alternation!
Comparison of the parallelism effects

But Conjoin Likes > genitive parallelism!

Roger Levy
The Internal Structure of Coordinate Categories
Grammar has very little to say about **categorical** constraints on the relation between conjuncts.
Discussion

- Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- But corpus data suggest there’s much more to say about the **gradient** constraints on their relation
Grammar has very little to say about **categorical** constraints on the relation between conjuncts.

But corpus data suggest there’s much more to say about the **gradient** constraints on their relation.

We now have the technical tools to formally characterize these gradient constraints.
Grammar has very little to say about **categorical** constraints on the relation between conjuncts

But corpus data suggest there’s much more to say about the **gradient** constraints on their relation

We now have the technical tools to formally characterize these gradient constraints

This formalization revealed a weakness of (sparse) corpus data and guided experiments to test for and quantify the strength of these constraints
Discussion

- Grammar has very little to say about **categorical** constraints on the relation between conjuncts.
- But corpus data suggest there’s much more to say about the **gradient** constraints on their relation.
- We now have the technical tools to formally characterize these gradient constraints.
- This formalization revealed a weakness of (sparse) corpus data and guided experiments to test for and quantify the strength of these constraints.
- We found that gradient “Conjoin Likes” is real, and has greater explanatory reach than was ever claimed for the categorical version!
Discussion

But why should gradient “Conjoin Likes” exist in the first place?
Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
- Psychological mechanism (priming)?
But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
- Psychological mechanism (priming)?
- Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?
Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
- Psychological mechanism (priming)?
- Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?
But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
- Psychological mechanism (priming)?
- Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?
But why should gradient “Conjoin Likes” exist in the first place?

- Stylistic preference?
- Psychological mechanism (priming)?
- Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?

- The Sag tradition of precise formal claims, serious engagement with data, and rigor in assigning credit for distributional generalizations will be essential to working this out.
Thank you, Ivan!

Figure 5: Double-set lattice ordered by $\supset$ over $\{A, B, C\}$. 


References II


Roger Levy | The Internal Structure of Coordinate Categories
References III


