Using the Arabic Treebank for Lexical Semantics

Neal Snider
Stanford University

Mona Diab
Columbia University
Levin Hypothesis (LH)

- Syntactic alternation behavior defines meaning
- Verbs that undergo similar alternations share aspects of meaning
- What aspects?
  - Event structure
Example

• Levin (1993):
  – break
  – hit
  – cut
  – touch
Transitive construction

<NP-SBJ NP-OBJ>
  – Margaret cut the bread
  – Janet broke the vase
  – Terry touched the cat.
  – Carla hit the door.
Middle construction

<NP-SBJ ADV>

– The bread cuts easily.
– Crystal vases break easily.
– *Cats touch easily.
– *Door frames hit easily.
Conative construction

<NP-SBJ PP-at>

– Margaret cut at the bread
– *Janet broke at the vase
– *Terry touched at the cat.
– Carla hit at the door.
Body part possessor

\(<NP-SBJ\ NP-OBJ>\ <NP-SBJ\ NP-OBJ\ PP-on>\)

- Margaret cut Bill’s arm
  Margaret cut Bill on the arm.
- Janet broke Bill’s finger.
  *Janet broke Bill on the finger.
- Terry touched Bill’s shoulder.
  *Terry touched Bill on the shoulder
- Carla hit Bill’s back.
  Carla hit Bill on the back
### Pattern of behavior

<table>
<thead>
<tr>
<th>Verb</th>
<th>Alt</th>
<th>Conative</th>
<th>Body part possessor</th>
<th>Middle</th>
</tr>
</thead>
<tbody>
<tr>
<td>touch</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>hit</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>cut</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>break</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
Verbs represent classes

- Break verbs:
  break, crack, rip, shatter, snap...
- Cut verbs:
  cut, hack, saw, scratch, slash...
- Touch verbs:
  pat, stroke, tickle, touch...
- Hit verbs:
  bash, hit, kick, pound, tap, whack...
What do the classes share?

• Event structure
  – States
  – Activities
  – Changes between the two

• Causation
cut verbs

• “bringing about a change of state by means of contact through motion”
• change of state with associated activity
  => achievement (Hale and Keyser, 1986)
break verbs

• pure change of state (without associated activity)
  => accomplishment
• Test: causative/inchoative alternation
  – The window broke.
    The little boy broke the window.
Cross-linguistic LH

- Warlpiri has conative alternation (Guerssel et al. 1985)
  - Also holds for hit, cut, but not break, touch
- No reason to expect alternations to hold cross-linguistically
- Arabic: no large scale study (but see Mahmoud, 1991)
Larger Project

- Create large scale verb classes for MSA
- Extract argument structure
- Use modern machine learning techniques to cluster verbs based on their argument structure(s)
- Inform computational and theoretical linguistics
Arabic Treebank

• Compiled by LDC
• 1800 stories of MSA newswire text from three different press agencies
• mostly politics, economics and sports journalism
• 800,000 Arabic tokens
• Parsed syntactically
Sample tree

( \( S \) (VP (PV+PVSUFF_SUBJ:3MS قال)
(NP-SBJ (NOUN_PROP دروري)
(S (ADVP-LOC (ADV ثمة)
(NP-TPC-1 (NOUN+CASE_INDEF_NOM ضرار(وقفت)
(VP (PV+PVSUFF_SUBJ:3FS وقعت)
(NP-SBJ-1 (-NONE- *T*))))))

\[
\begin{align*}
S & \quad | \quad VP \\
V & \quad | \quad NP \\
قال & \quad | \quad دروري & \quad | \quad ثمة & \quad | \quad ضرار & \quad | \quad وقعت
\end{align*}
\]
Extracting alternations

• NP arguments annotated (NP-SBJ, NP-OBJ, etc.)
• PPs deemed essential to verb meaning annotated PP-CLR
• Assume these are also arguments
Extracting alternations

- Create argument structure matrix:

<table>
<thead>
<tr>
<th>Verb</th>
<th>Args</th>
<th>&lt;NP-SBJ, NP-OBJ&gt;</th>
<th>...</th>
<th>&lt;NP-SBJ PP-CLR-علي&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>لغارة</td>
<td>3</td>
<td>...</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>تمزق</td>
<td>4</td>
<td>...</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

- Possible alternations = intersection of verbs that are found in >2 columns
Alternation

- فـي

- <NP-SBJ PP-CLR-بـفـي>, <NP-SBJ PP-CLR-في>

سببت الامطار الغزيرة في ارتفاع منسوب المياه

tatasab~abu fiy Almaziydi min safoki AldmA
The abundant rain caused a rise in water levels (internal cause)

تسببت بـمقتل 11 شخصا

tasab~abat bi-maqotali 11 $axoSAF
It caused the killing of 11 people (external cause)

- تحكم

(taHak~am, control)

تسبب
(tasab~ab, cause)
Alternation على SBAR

• <NP-SBJ NP-OBJ SBAR> <NP-SBJ NP-OBJ PP-CLR- على>

حث-هم أن يتوجهوا ب- نية التفكركي خلق الله
Hav-a-hum >an yatawaj-ahuwA bi-niy-api Altafak-uri fiy xaloqi All~`h
He urges them to think of God’s creation (achievement)

حث شافيز دول منظمة اوبيك على مواجهة ضغوطات
Hav-a $Afiyz duwala munaZ-amapi >uwbik EalaY muwAjahapi DuguwTAti
Chavez urges the OPEC countries to confront the pressures (activity)

• حتم (Hat-am, prescribe/make a duty/decide)

سلط (sal-aT, impose/exert)

حث (Hav-, encourage/urge/incite)

أرغم (>arogam, compel/coerce/force)?

أجب (ajobar, force/compel)?
Alternation ب-drop

- <NP-SBJ NP-OBJ> <NP-SBJ NP-OBJ PP-CLR- ب>

أحاط مئات من رجال الشرطة الأزهر
>aHATa mi}AtN min rijAli Al$uroTapi Al>azohara
Hundreds of police surrounded the Azhar (achievement)

أحيط مكان-ها ب-سرية كبيرة
uHiyTa makAnu <ijorA}i-hA bi-sir~iy~apK kabiyrapK
Its location was engulfed in great secrecy (activity)

• بت (bat~, determine/complete)
• أحاط (>aHAT, surround/encircle)
Conclusions

• Possible alternations in Arabic
• Corpora can complement linguists’ intuitions in formulating hypotheses
• Classes are evidence that Levin Hypothesis holds cross linguistically