Frequent frames in CDS as training wheels for acquisition of the verb paradigm in Hebrew

Tamar Johnson and Naomi Havron

Can frequent frames in Child Directed Speech (CDS) assist children in acquiring the complex structure of the verb paradigm in Hebrew? Hebrew, as a Semitic language, is featured with non-linear morphology. Verbs are made up of roots or radical elements (the *shoresh* and *binyanim*). The Hebrew verb paradigm is built of seven possible binyanim, in three tenses and two genders. Despite the complexity of the system, children do more than recognize the more conspicuous and reliable roots radicals, but are actually recognizing that the same root can take several binyanim. In input to children, however, around 70% of all verbs are singletons: each root appearing with only one binyan. Only about 25% of verbs appear in 2-binyan families [1]. With such input, how do children accomplish learning about binyanim, given that overwhelming variety of forms verbs can take on?

One property which characterizes CDS, and probably makes it easier for children to learn, is a large amount of repetition [2-6]. Specifically, CDS contains many frequent frames or lexically restricted utterances. Lexically restricted utterances have been found to promote linguistic processing, such that children make less mistakes when they use lexically restricted utterances [7,8], and respond faster to them [9,10].

Here, we examined verbs which followed a frequent frame in CDS taken from the Childes corpus in Hebrew. The verbs outside the boundaries of the frequent frame can vary more than the frame itself, but because of their appearance right after a frequent frame, they have common elements introduced by the frame. We predict that:

1. The frequent frame creates a set of verbs that can appear right after it. This set is smaller than the general group of verbs in the lexicon. Thus, we predict the entropy of the word following the frequent frame to be lower than the general entropy of words in the lexicon to such a level of entropy, which can be more accessible for children to process.
2. Each frequent frame creates a slightly different probability distribution of the properties the following verb possesses. i.e. Some properties of verbs have a high probability of appearing after a specific frequent frame, and other properties have a low probability of appearing in that location. This information, gathered from the child's experience, can help him attribute information to the morphological properties of the verb.

As a preliminary step, we examined verbs which followed the three most frequent bigrams in Hebrew CDS [11] and as predicted, the frames differed in the distributions of binyan, tense, and root the following verb possessed (see attached appendix). In addition, the distribution of verbs following the three frequent frames by binyan also differed from the distribution in the general lexicon [1]. These findings strengthens the hypothesis that each frequent frame reduces the complexity of the verb paradigm into a more limited set of verbs that may show a level of variation ideal for learning conditions, meaning they have a moderate level of new information vs. predicted elements, termed a *Goldilocks effect* [12]. An in-progress and planned work is testing our predictions exploring a wider range of frequent frames using computational methods.
Appendix: results

Table 1: Distribution of verbs after the three frequent frames by tenses:

<table>
<thead>
<tr>
<th></th>
<th>Ma at</th>
<th>Ani lo</th>
<th>At lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past</td>
<td>1.9%</td>
<td>5.1%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Present</td>
<td>97.3%</td>
<td>84.3%</td>
<td>90.3%</td>
</tr>
<tr>
<td>Future</td>
<td>0.7%</td>
<td>10.6%</td>
<td>6.1%</td>
</tr>
</tbody>
</table>

*Out of cases where verbs followed the frames

Table 2: Distribution of verbs after the three frequent frames by binyan:

<table>
<thead>
<tr>
<th></th>
<th>Ma at</th>
<th>Ani lo</th>
<th>At lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kal</td>
<td>86.1%</td>
<td>61%</td>
<td>85.9%</td>
</tr>
<tr>
<td>Piel</td>
<td>8.5%</td>
<td>1.3%</td>
<td>5.8%</td>
</tr>
<tr>
<td>Hif'il</td>
<td>1.9%</td>
<td>34.3%</td>
<td>6.45%</td>
</tr>
<tr>
<td>Hitpa'el</td>
<td>2.8%</td>
<td>0.8%</td>
<td>0.88%</td>
</tr>
</tbody>
</table>

*The verb Hyit (תָּיִיה) was counted as verb in past tense but does not have binyan [13]. That is the reason that the numbers do not to add up to a 100 percent in each row.

References:

[1] Ravid et. al, 2015;
[5] Lieven, Salomo, & Tomasello, 2009;
[7] Ambridge, Rowland, Theakston, & Tomasello, 2006;
[10] Fernald & Hurtado, 2006;
[13] Itai & Wintner, 2008;