Psych 215L: Language Acquisition

Lecture 12
Morphosyntax

Computational Problem

Determine that there are grammatical categories like Noun and Verb that behave similarly with respect to their morphology and combinatorial syntax.

Noun = {penguin, goblin, glitter, cheese}
Morphology: Nouns can take determiners like "the"
{the penguin, the goblin, the glitter, the cheese}

Verb = {swim, dance, flutter, smell}
Morphology: Verbs can take –ed to indicate past tense
Combinatorial syntax: Verbs can take adverbs that modify them, like "really"
{really swim, really dance, really flutter, really smell}

Yang 2010

How do we know when children achieve adult-like knowledge?

"Language use is the composite of linguistic, cognitive and perceptual factors many of which, in the child's case, are still in development and maturation. It is therefore difficult to draw inferences about the learner's linguistic knowledge from his linguistic behavior."

"The pioneering work on child language that soon followed, include those who did not follow the generative approach, also recognized the gap between what the child knows and what the child says... child language be interpreted in terms of adult-like grammatical devices, which has continued to feature prominently in language acquisition."

Example adult-like grammatical device: Verb categories like Noun and Verb

Yang 2010

How do we know when children achieve adult-like knowledge?

"This tradition has been challenged by the *item* or *usage-based* approach to language most clearly represented by Tomasello (1992, 2000a, 2000b, 2003), which reflects a current trend (Bybee 2001, Pierre- humbert 2001, Goldberg 2003, Culicover & Jackendoff 2005, Hay & Baayen 2005, etc.) that emphasizes the storage of specific linguistic forms and constructions at the expense of general combinatorial linguistic principles and overarching points of language variation (Chomsky 1965, 1981)."

Properties used in support of item-based approach:
(1) Use of verb in limited "constructions"
(2) Limited morphology on any given verb
(3) Unbalanced determiner usage (ex: use only "the" with some and only "a/an" with others)

Yang 2010

The lack of a formal statistical test for productivity

"So far as we can tell, however, these evidence in support for item-based learning has been presented, and accepted, on the basis of intuitive inspections rather than formal empirical tests. For instance, among the numerous examples from child language, no statistical test was given in the major treatment (Tomasello 1992) where the Verb Island Hypothesis and related ideas about item-based learning are put forward. Specifically, no test has been given to show that the observations above are statistically *inconsistent* with the expectation of a fully productive grammar, the position that item-based learning opposes. Nor, for that matter, are these observations shown to be *consistent* with item-based learning,…"

Yang 2010

Zipf's law

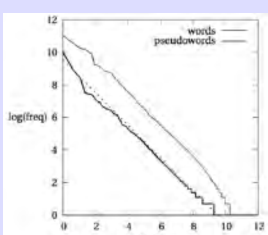
"Under the so-called *Zipf 's law* (Zipf 1949), the empirical distributions of words follow a curious pattern: relatively few words are used frequently—very frequently—while most words occur rarely, with many occurring only once in even large samples of texts. More precisely, the frequency of a word tends to be approximately inversely proportional to its rank in frequency."

f = frequency
r = rank

$f = \frac{C}{r}$ where C is some constant

Yang 2010

Checking Zipf's law on the Brown corpus



"The lower line is plotted by taking "words" to be any sequence of letters between *e's* (Chomsky 1958). The two straight dotted lines are linear functions with the slope -1, which illustrate the goodness of the Zipfian fit."

Yang 2010

Checking Zipf's law on the Brown corpus

"It is often the case that we are not concerned with the actual frequencies of words but their probability of occurrence; Zipf's law makes this estimation simple and accurate."

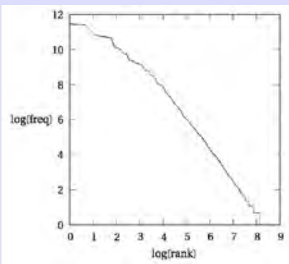
$$p_r = \left(\frac{C}{r}\right) \Big/ \left(\sum_{i=1}^{N} \frac{C}{i}\right) = \frac{1}{rH_N} \text{ where } H_N \text{ is the } N\text{th Harmonic Number } \sum_{i=1}^{N} \frac{1}{i}$$

r = rank of word
p_r = probability of the occurrence of word with rank r
N = number of word types in corpus

Basic intuition:
p_x = [frequency of x]/[total frequency of all items]

Yang 2010

Checking Zipf's law on the syntactic rules in the Penn-Treebank corpus



"Since the corpus has been manually annotated with syntactic structures, it is straightforward to extract rules and tally their frequencies. The most frequent rule is "PP→P NP", followed by "S→NP VP": again, the Zipf-like pattern."

Yang 2010

The moral of Zipf's law for productivity analyses

"Claims of item-based learning build on the premise that linguistic productivity entails diversity of usage: the "unevenness" in usage distribution is taken to be evidence against a systematic grammar. The underlying intuition, therefore, appears to be that linguistic combinations might follow something close to a uniform distribution."

A closer look at determiner usage with nouns (among other types of usage)
"Consider a fully productive rule that combines a determiner and a singular noun, or "DP→ D N", where "D→ a|the" and "N→ cat|book|desk|...". We use this rule for its simplicity and for the readily available data for empirical tests but one can easily substitute the rule for "VP→ V DP", "VP→ V in Construction$_x$ ", "$V_{inflection} \rightarrow V_{stem}$ + Person + Number + Tense". All such cases can be analyzed with the methods provided here."

Yang 2010

Expected determiner usage

"Suppose a linguistic sample contains *S* determiner-noun pairs, which consist of *D* and *N* unique determiners and nouns. (In the present case *D* = 2 for "a" and "the".) The full productivity of the DP rule, by definition, means that the two categories combine independently."

Observation 1
"…nouns (and open class words in general) will follow Zipf's law…relatively few nouns occur often but many will occur only once—which of course cannot overlap with more than one determiners."

Observation 2
"…while the combination of *D* and *N* is syntactically interchangeable, *N*'s tend to favor one of the two determiners, a consequence of pragmatics and indeed non-linguistic factors."

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Overlap: A noun occurs with more than one determiner.

Calculating observed overlap
For each noun n in the data set, determine if it occurs with more than one determiner.
If so, overlap(n) = 1.
If not, overlap(n) = 0.

Observed overlap = $\dfrac{\sum_{N} overlap(n)}{N}$

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

"This requires the calculation of the expected overlap value for each of the *N* nouns over all possible compositions of the sample."

$$O(D,N,S) = \frac{1}{N}\sum_{r=1}^{N} O(r,N,D,S)$$

Sum individual expected overlap for each noun (from rank 1 to N) in the data set, and then divide by N to get the average expected overlap for all nouns.

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Individual noun overlap: Probability that it is not used with only one determiner.

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Individual noun overlap: Probability that it is not used with only one determiner.

All the instances where it's not the case that…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Individual noun overlap: Probability that it is not used with only one determiner.

…the noun just didn't get produced in this data set of S samples for whatever reason…

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1-(1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Individual noun overlap: Probability that it is not used with only one determiner.

…and the noun was sampled, but favored one determiner exclusively for whatever reason.

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1-(1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun just didn't get produced for S samples

($1-p_r$) = probability that noun with rank r didn't appear for this one trial

…done S times (quantity^S)

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1-(1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(1) probability of noun n_r (which appears with frequency p_r) combining with the i^{th} determiner (which has its own frequency of appearing in the corpus, d_i) = p_r*d_i = $d_i p_r$

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1-(1-p_r)^S \\ &\quad - \sum_{i=1}^{D}\left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + \ldots + p_{r-1} + d_i p_r + p_{r+1} + \ldots + p_N)^S$$

However frequently noun with rank 1 appeared with whatever determiners +…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + ... + p_{r-1} + d_i p_r + p_{r+1} + ... + p_N)^S$$

…however frequently noun with rank 2 appeared with whatever determiners +…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + ... + p_{r-1} + d_i p_r + p_{r+1} + ... + p_N)^S$$

…however frequently noun with rank r-1 appeared with whatever determiners +…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + ... + p_{r-1} + d_i p_r + p_{r+1} + ... + p_N)^S$$

…how frequently this noun with rank r appeared and with only this one determiner d_i +…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\ &\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + ... + p_{r-1} + d_i p_r + p_{r+1} + ... + p_N)^S$$

…and so on…

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$
\begin{aligned}
O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\
&\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\
&= 1 - (1-p_r)^S \\
&\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right]
\end{aligned}
$$

Calculating the probability that this noun favored one determiner exclusively

(2) probability of all possible compositions of sample S where n_r combines with d_i only

$$(p_1 + p_2 + \ldots + p_{r-1} + d_i p_r + p_{r+1} + \ldots + p_N)^S$$

…for each of the S samples in the data set

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$
\begin{aligned}
O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\
&\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\
&= 1 - (1-p_r)^S \\
&\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right]
\end{aligned}
$$

Calculating the probability that this noun favored one determiner exclusively

(2) Since $(p_1 + p_2 + p_{r-1} + p_r + p_{r+1} + \ldots + p_N) = 1$

$$(p_1 + p_2 + \ldots + p_{r-1} + d_i p_r + p_{r+1} + \ldots + p_N)^S = (d_i p_r + 1 - p_r)^S$$

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$
\begin{aligned}
O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\
&\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\
&= 1 - (1-p_r)^S \\
&\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right]
\end{aligned}
$$

Calculating the probability that this noun favored one determiner exclusively

(2) Another way to derive $(d_i p_r + 1 – p_r)^S$

For each sample (quantity^S), we want the probability of *not* {picking that noun out when it doesn't come with determiner d_i}. This is $(1 - p_r(1 - d_i)) = 1 – p_r + d_i p_r = d_i p_r + 1 – p_r$.

Yang 2010

Quantifying productivity

S = # of samples in linguistic data set

D = # of unique determiners

N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$
\begin{aligned}
O(r,N,D,S) &= 1 - \Pr\{n_r \text{ is not sampled during } S \text{ trials}\} \\
&\quad - \sum_{i=1}^{D} \Pr\{n_r \text{ is sampled but with the } i\text{th determiner exclusively}\} \\
&= 1 - (1-p_r)^S \\
&\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S\right]
\end{aligned}
$$

Calculating the probability that this noun favored one determiner exclusively

(2) A third way to derive $(d_i p_r + 1 – p_r)^S$

For each sample (quantity^S), we can either pick out that noun with determiner d_i or we can pick some other noun besides n_r. This is $(d_i p_r + (1-p_r)) = d_i p_r + 1 – p_r$.

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S \right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

(3) However $(d_i p_r + 1 – p_r)^S$ includes the probability of n_r combining with d_i 0 times. We can especially see this under the last view of how to derive $d_i p_r + 1 – p_r$. For each sample, either we pick that noun with d_i, or we don't pick that noun. But this means that this quantity includes the probability that *for all S samples*, we didn't pick that noun = $(1-p_r)^S$. We need to subtract that off.

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$\begin{aligned} O(r,N,D,S) &= 1 - \Pr[n_r \text{ is not sampled during } S \text{ trials}] \\ &\quad - \sum_{i=1}^{D} \Pr[n_r \text{ is sampled but with the } i\text{th determiner exclusively}] \\ &= 1 - (1-p_r)^S \\ &\quad - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S - (1-p_r)^S \right] \end{aligned}$$

Calculating the probability that this noun favored one determiner exclusively

Note: This quantity is also equivalent to the following equation, which calculates $p(n_r$ is sampled but with d_i exclusively) directly:

For each combination of S samples… $\displaystyle\sum_{j=1}^{S} \binom{S}{j} \underbrace{(d_i * p_r)^j (1 - p_r)^{S-j}}$

All the permutations that have j uses of n_r

A sample where there are j uses of n_r and S-j uses of some other noun

Yang 2010

Quantifying productivity
S = # of samples in linguistic data set
D = # of unique determiners
N = # of unique nouns

Calculating expected overlap [O(D,N,S)]

$$O(r,N,D,S) = 1 + (D-1)(1-p_r)^S - \sum_{i=1}^{D} \left[(d_i p_r + 1 - p_r)^S \right]$$

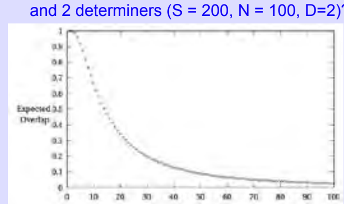
Collecting the terms together…

…and this is what we use in the original formula

$$O(D,N,S) = \frac{1}{N} \sum_{r=1}^{N} O(r,N,D,S)$$

Yang 2010

What kind of overlap do we expect in a sample of size 200 with 100 nouns, and 2 determiners (S = 200, N = 100, D=2)?



"As can be seen, few of nouns have high probabilities of occurring with both determiners, but most are (far) below chance. The average overlap is 21.1%."

Yang 2010

Determiner usage case study: D = {*a*, *the*} only

Data = Adam [Brown], Eve [Brown], Sarah [Brown], Naomi [Sachs], Peter [Bloom?], Nina [Suppes]

Age range across all children: 1;1 – 5;1

Comparison sets

Each individual child

+

First 100, 300, and 500 productions from all children to capture earliest stage of language production which should (presumably) be the least productive

vs.

Adult production estimates from the Brown corpus

Yang 2010

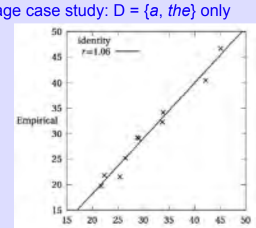
Determiner usage case study: D = {*a*, *the*} only

Subject	Sample Size (*S*)	*a* or *the* Noun types (*N*)	Overlap (expected)	Overlap (empirical)	$\frac{S}{N}$
Naomi (1;1-5;1)	884	349	21.8	19.8	2.53
Eve (1;6-2;3)	831	283	25.4	21.6	2.94
Sarah (2;3-5;1)	2453	640	28.8	29.2	3.83
Adam (2;3-4;10)	3729	780	33.7	32.3	4.78
Peter (1;4-2;10)	2873	480	42.2	40.4	5.99
Nina (1;11-3;11)	4542	660	45.1	46.7	6.88
First 100	600	243	22.4	21.8	2.47
First 300	1800	483	29.1	29.1	3.73
First 500	3000	640	33.9	34.2	4.68
Brown corpus	20650	4664	26.5	25.2	4.43

"The theoretical expectations and the empirical measures of overlap agree extremely well.... Neither paired t- nor Wilcoxon test reveal significant difference between the two sets of values."

Yang 2010

Determiner usage case study: D = {*a*, *the*} only



"Perhaps a more revealing test is linear regression (Figure 5): a perfect agreement between the two sets of value would have the slope of 1.0, and the actual slope is 1.08 (adjusted R^2 = 0.9716). Therefore, we could that the determiner usage data from child language is consistent with the productive rule "DP→ D N"."

Yang 2010

Determiner usage case study: D = {*a*, *the*} only

Subject	Sample Size (*S*)	*a* or *the* Noun types (*N*)	Overlap (expected)	Overlap (empirical)	$\frac{S}{N}$
Naomi (1;1-5;1)	884	349	21.8	19.8	2.53
Eve (1;6-2;3)	831	283	25.4	21.6	2.94
Sarah (2;3-5;1)	2453	640	28.8	29.2	3.83
Adam (2;3-4;10)	3729	780	33.7	32.3	4.78
Peter (1;4-2;10)	2873	480	42.2	40.4	5.99
Nina (1;11-3;11)	4542	660	45.1	46.7	6.88
First 100	600	243	22.4	21.8	2.47
First 300	1800	483	29.1	29.1	3.73
First 500	3000	640	33.9	34.2	4.68
Brown corpus	20650	4664	26.5	25.2	4.43

"Given *N* unique nouns in a sample of *S*, [a] greater overlap value can be obtained if more nouns occur more than once. That is, words whose probabilities are greater than 1/S can increase the overlap value."

Yang 2010

Determiner usage case study: D = {*a*, *the*} only

Subject	Sample Size (S)	*a* or *the* Noun types (N)	Overlap (expected)	Overlap (empirical)	$\frac{S}{N}$
Naomi (1;1-5;1)	884	349	21.8	19.8	2.53
Eve (1;6-2;3)	831	283	25.4	21.6	2.94
Sarah (2;3-5;1)	2453	640	28.8	29.2	3.83
Adam (2;3-4;10)	3729	780	33.7	32.3	4.78
Peter (1;4-2;10)	2873	480	42.2	40.4	5.99
Nina (1;11-3;11)	4542	660	45.1	46.7	6.88
First 100	600	243	22.4	21.8	2.47
First 300	1800	483	29.1	29.1	3.73
First 500	3000	640	33.9	34.2	4.68
Brown corpus	20650	4664	26.5	25.2	4.43

"Zipf's law…allows us to express this cutoff line in terms with ranks, as the probability of the noun n_r with rank r [= p_r] has the probability of $1/(r*H_N)$. The derivation…uses the fact that the Nth Harmonic Number $\sum_{i=1}^{N} 1/i$ can be approximated by ln N ."

$$S\frac{1}{rH_N}=1$$
$$r=\frac{S}{H_N}\approx\frac{S}{\ln N}$$

Yang 2010

Determiner usage case study: D = {*a*, *the*} only

Subject	Sample Size (S)	*a* or *the* Noun types (N)	Overlap (expected)	Overlap (empirical)	$\frac{S}{N}$
Naomi (1;1-5;1)	884	349	21.8	19.8	2.53
Eve (1;6-2;3)	831	283	25.4	21.6	2.94
Sarah (2;3-5;1)	2453	640	28.8	29.2	3.83
Adam (2;3-4;10)	3729	780	33.7	32.3	4.78
Peter (1;4-2;10)	2873	480	42.2	40.4	5.99
Nina (1;11-3;11)	4542	660	45.1	46.7	6.88
First 100	600	243	22.4	21.8	2.47
First 300	1800	483	29.1	29.1	3.73
First 500	3000	640	33.9	34.2	4.68
Brown corpus	20650	4664	26.5	25.2	4.43

So for Naomi, we expect only the first 2 or 3 ranked nouns to have a non-zero overlap.

For the Brown corpus, we expect only the first 4 or 5 ranked nouns to have a non-zero overlap.

Yang 2010

How do we evaluate the item-based approach, though?

"In the limiting case, the item-based child learner could store the input data in its entirety and simply retrieve these memorized determiner-noun pairs in production. Since the input data, which comes from adults, is presumably productive, children's repetition of it may show the same degree of productivity."

"Tomasello (2000c, p77) suggests that "…so they simply retrieve that expression from their stored linguistic experience." Following this line of reasoning, we consider a learning model that memorizes *jointly* formed, as opposed to productively composed, determiner-noun pairs from the input; presumably these "stored linguistic experience" as such nouns (and determiners) constitute a large part of adult-child linguistic communication in every- day life. These pairs will then be sampled directly…"

Yang 2010

How do we evaluate the item-based approach, though?

global memory learner: composite of all children's input

local memory learner: drawn just from one particular child's input

"For each child, then, there are two sets of data: the determiner-noun pairs along with their frequencies from that child's input (local memory learner) and the determiner-noun pairs along with their frequencies in the entire 1.1 million utterances of adult speech (global memory learner)....we use the Monte Carlo simulation to randomly draw *S* pairs from the two sets of data that correspond to the local and global memory learning models. The probability with which a pair is drawn is proportional to its frequency in the two sets of data.…We calculate the value of overlap from this list, that is, the percentage of nouns that appear with both "a" and "the" over the total number of nouns. The results are averaged over 1000 draws."

Yang 2010

How do we evaluate the item-based approach, though?

global memory learner: composite of all children's input

local memory learner: drawn just from one particular child's input

Child	Sample Size (S)	Overlap (global memory)	Overlap (local memory)	Overlap (empirical)
Eve	831	16.0	17.8	21.6
Naomi	884	16.6	18.9	19.8
Sarah	2453	24.5	27.0	29.2
Peter	2873	25.6	28.8	40.4
Adam	3729	27.5	28.5	32.3
Nina	4542	28.6	41.1	46.7
First 100	600	13.7	17.2	21.8
First 300	1800	22.1	25.6	29.1
First 500	3000	25.9	30.2	34.2

"Both sets of overlap values from the two variants of item-based learning…differ significantly from the empirical measures: $p < 0.005$ for both paired t-test and paired Wilcoxon test. This suggests that children's use of determiners does not follow the predictions of the item-based learning approach…"

Yang 2010

How do we evaluate the item-based approach, though?

global memory learner: composite of all children's input

local memory learner: drawn just from one particular child's input

Child	Sample Size (S)	Overlap (global memory)	Overlap (local memory)	Overlap (empirical)
Eve	831	16.0	17.8	21.6
Naomi	884	16.6	18.9	19.8
Sarah	2453	24.5	27.0	29.2
Peter	2873	25.6	28.8	40.4
Adam	3729	27.5	28.5	32.3
Nina	4542	28.6	41.1	46.7
First 100	600	13.7	17.2	21.8
First 300	1800	22.1	25.6	29.1
First 500	3000	25.9	30.2	34.2

"Naturally, our evaluation here is tentative since the proper test can be carried out only when the theoretical predictions of item-based learning are made clear. And that is exactly the point: the advocates of item-based learning not only rejected the alternative hypothesis without adequate statistical tests, but also accepted the favored hypothesis without adequate statistical tests."

Yang 2010

Case study: Verbal morphology

"Few stems appear in a great number of inflections, which, however, never approach anywhere near the maximum number of possible inflections. Moreover, most stems are used very sparsely, the majority of which occur in exactly one inflection."

Example: Spanish verb morphology [1st, 2nd, 3rd person + sg vs. pl]

	present tense, imperfect aspect, indicative mood, -ar verb	past tense, perfect aspect, indicative mood, -ar verb	present tense, imperfect aspect, subjunctive mood, -ar verb	past tense, imperfect aspect, indicative mood -ar verb	…
1s	hablo	hablé	hable	hablaba	
2s	hablas	hablaste	hables	hablabas	
3s	habla	habló	hable	hablaba	
1p	hablamos	hablamos	hablemos	hablábamos	
2p	habláis	hablasteis	habléis	hablabais	
3p	hablan	hablaron	hablen	hablaban	

Yang 2010

Case study: Verbal morphology

Survey of inflectional usage data in Italian, Spanish, and Catalan

6 forms = 1st, 2nd, & 3rd person + sg vs. pl

Subjects	1 form	2 forms	3 forms	4 forms	5 forms	6 forms	S/N
Italian children	81.8	7.7	4.0	2.5	1.7	0.3	1.533
Italian adults	63.9	11.0	7.3	5.5	3.6	2.3	2.544
Spanish children	80.1	5.8	3.9	3.2	3.0	1.9	2.233
Spanish adults	76.6	5.8	4.6	3.6	3.3	3.2	2.607
Catalan children	69.2	8.1	7.6	4.6	3.8	2.0	2.098
Catalan adults	72.5	7.0	3.9	4.6	4.9	3.3	2.342

"…the logic of the problem remains the same…the diversity of usage depends on the number of opportunities for a verb stem to appear multiple forms, or *S/N* ….children learning Spanish and Catalan show very similar agreement usage to adults—and the *S/N* ratios are also very similar for these groups."

Yang 2010

Case study: Verbal morphology

Survey of inflectional usage data in Italian, Spanish, and Catalan

6 forms = 1st, 2nd, & 3rd person + sg vs. pl

Subjects	1 form	2 forms	3 forms	4 forms	5 forms	6 forms	S/N
Italian children	81.8	7.7	4.0	2.5	1.7	0.3	1.533
Italian adults	63.9	11.0	7.3	5.5	3.6	2.3	2.544
Spanish children	80.1	5.8	3.9	3.2	3.0	1.9	2.233
Spanish adults	76.6	5.8	4.6	3.6	3.3	3.2	2.607
Catalan children	69.2	8.1	7.6	4.6	3.8	2.0	2.098
Catalan adults	72.5	7.0	3.9	4.6	4.9	3.3	2.342

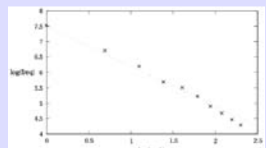
"Italian children use somewhat more stems in only one form than Italian adults (81.8% vs. 63.9%), but that follows from the *S/N* ratio (2.544 vs. 1.533). That is, for each verb, the Italian adults have roughly 66% more opportunities to use it than the Italian children, which would account for the discrepancy in the frequency of one-form verbs."

Yang 2010

Case study: Verb arguments

"We focus on constructions that involve a transitive verb and its nominal objects, including pronouns and noun phrases. Following the definition of "sentence frame" in Tomasello's original Verb Island study (1992, p242), each unique lexical item in the object position counts as a unique construction for the verb."

Zipfian distribution for top 15 transitive verbs from 1.1 million utterances of child-directed speech



"…even for large corpora, a verb appears in few constructions frequently and in most constructions infrequently if at all. The observation of Verb Islands, that verbs tend to combine with one or few elements out of a large range, is in fact characteristic of a fully productive verbal syntax system."

Kowalski & Yang 2012

Case study: Verb arguments

"For each verb, we count the frequencies of its top 10 most frequent constructions, which are defined as the verb followed a unique lexical item in the object position (e.g., "ask him" and "ask John" are different constructions, following Tomasello 1992)."

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
put	401	164	124	15	12	12	11	10	8	5
tell	245	64	49	49	45	36	22	16	14	13
see	152	100	38	32	28	21	14	14	12	11
want	158	83	36	24	19	15	13	9	5	4
let	238	38	32	23	22	17	8	6	3	3
give	115	92	59	32	31	7	5	5	5	5
take	130	57	30	21	18	15	14	9	8	7
show	100	34	27	21	19	17	12	8	7	7
got	58	37	14	12	11	9	7	7	7	4
ask	45	41	27	24	12	10	8	8	4	2
make	67	20	12	10	9	7	7	4	3	2
eat	67	42	14	8	6	5	5	3	3	3
like	39	13	9	6	4	4	4	4	3	3
bring	43	30	17	15	10	10	3	3	3	3
hear	46	22	13	9	6	4	4	3	3	3
total	1904	838	501	301	252	189	137	109	88	75

Yang 2010

Case study: Verb arguments

How many samples would we need to see in order to see verbs combining with 50% of the objects they could combine with?

Vocabulary: 100 verbs, 100 potential objects [10,000 combinations]

→ Monte Carlo sampling simulation: ~28,000 samples

→ Approximate amount of production data: 9.6 million words

Vocabulary: 1500 verbs, 1500 potential objects [2,250,000 combinations]

→ Monte Carlo sampling simulation: ~1.4 million samples

→ Approximate amount of production data: ~4.8 billion words (46 years of non-stop talking)

Basic point:

Unlikely to ever see anything *except* verb islands in production data

Yang 2010

Take home points

"For any type of linguistic expression that involve open class items—and that means *every* type of linguistic expression—modest measures of usage diversity requires extremely large samples."

"Zipf's law hints at the inherent limitations in approaches that stress the storage of construction-specific rules or processes…the Zipfian distribution of linguistic combinations…ensure that most "pairings of form and function" simply will never be heard, never mind stored, and those that do appear may do so with sufficiently low frequency such that no reliable storage and use is possible."

Yang 2010

Take home points

"The sparse data problem strikes…and the role of memory in language learning should not be overestimated. In linguistics and cognitive science, of course, the learner's Zipfian challenge bears another name: the argument from the poverty of stimulus…To attain full linguistic competence, the child learner must overcome the Zipfian distribution and draw generalizations about language on the basis of few and narrow types of linguistic expressions."