Word learning and translation

- Word learning and translation are analogous
- Ostension = "definition by showing"
- Word - meaning pairs
- What is referred to by a given word is often ambiguous (indeterminacy of translation)

  - Scenes and sentences are complex and need to be parsed

  - Subordinate/superordinate status is always ambiguous

  - Philosophical problems, e.g. "Gavagai" (Quine)

Eskimo words for snow

- Supposedly, the Eskimos (Yupik/Inuit) have many words for SNOW

  - This turns out not be true but is still an interesting intuition pump!

- Sapir-Whorf hypothesis: Language influences thought , aka Linguistic Relativity (see Hopi)

  Whorf: Eskimos think about snow differently because they have different snow words from us!

  Modern perspective: Eskimos know more about snow, and make distinctions we don't make, so naturally they have more basic-level words for kinds of snow (cf: horse, cook, gun…)

  (Thought/experience influences language)

  Modern perspective: Many Eskimo words for snow turn out to be morphologically decomposable, e.g. "soft-snow" "wet-snow", etc.

  By that standard we have many words for snow too, e.g. "soft snow", "wet snow", etc.

Coding efficiency

- Shannon (1959) showed how to quantify information (bits, bytes, etc.)
- In order to achieve the most efficient encoding of a set of concepts, assign the shortest codes to the most frequent concepts
- If you have a compositional code, this means that high-frequency concepts become primitive symbols, and lower-frequency concepts are expressed by combinations of primitive symbols
- For example, assign individual words to very frequently encountered concepts (lexicalization), and multi-word phrases to superordinate concepts and subordinate concepts

Morse code

Letter	Morse	Letter	Morse	Letter	Morse
A	.-	N	-.	0	-----
B	-...	O	---	1	.----
C	-.-.	P	.--.	2	..---
D	-..	Q	--.-	3	...--
E	.	R	.-.	4	....-
F	..-.	S	...	5	.....
G	--.	T	-	6	-....
H	....	U	..-	7	--...
I	..	V	...-	8	---..
J	.---	W	.--	9	----.
K	-.-	X	-..-		
L	.-..	Y	-.--		
M	--	Z	--..		

Morse code obeys Shannon's principle: the more frequent the symbol, the shorter its code

Letter	Code	Length	Frequency
E	.	1	12.49%
T	-	3	9.28%
A	.-	4	8.04%
O	---	9	7.64%
I	..	2	7.57%
N	-.	4	7.23%
S	...	3	6.51%
R	.-.	5	6.28%
H	....	4	5.05%
L	.-..	6	4.07%
D	-..	5	3.82%
C	-.-.	8	3.34%
U	..-	5	2.73%
M	--	6	2.51%
F	..-.	6	2.40%
P	.--.	8	2.14%
G	--.	7	1.87%
W	.--	7	1.68%
Y	-.--	10	1.66%
B	-...	6	1.48%
V	...-	6	1.05%
K	-.-	7	0.54%
X	-..-	8	0.23%
J	.---	10	0.16%
Q	--.-	10	0.12%
Z	--..	8	0.09%

Lexicalization

- The result is that expertise in a domain area leads to the recoding of complex concepts into basic-level concepts, i.e. lexicalization

  - Horse experts call horses mares, stallions, geldings, etc.

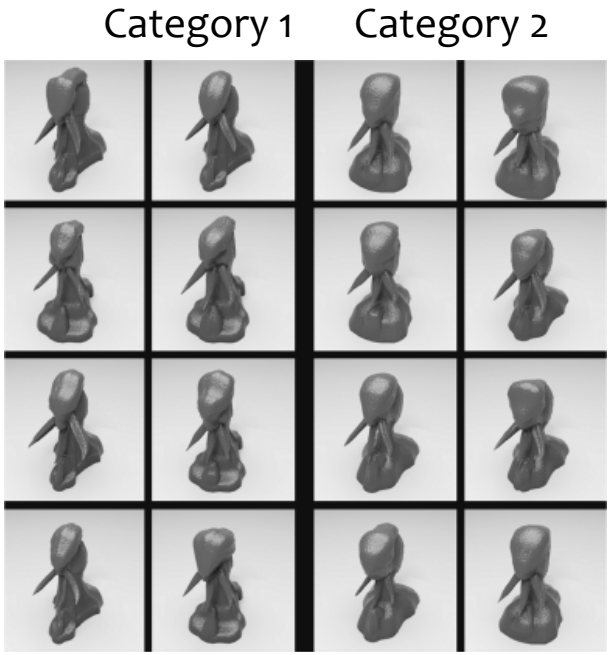
  - Novices *cook* and *bake*. Experts *boil*, *roast*, *braise*, *sear*, *parbioil*, et.

- Cultural differences can be seen as differences in expertise

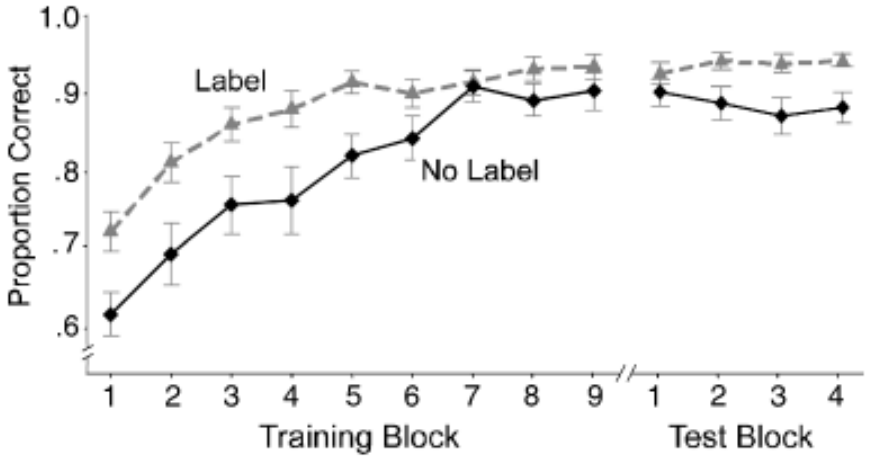
- But notice this does not entail qualitative differences in concepts—just a reassignment of code lengths

Linguistic labels aid categorization

- Lupyan et al. (2007) trained subjects on two novel categories, either
  (a) with verbal labels
  (b) without verbal labels
- Verbal labels aid learning even when they provide no information



Stimuli



Human performance

- If words aid categorization, can they induce categorical perception?