LING5702: Lecture Notes 1
Introduction and Background

Contents

1.1 What is this course about?

This course will cover fundamental questions about what language is.

This course differs from other *psychology* courses because:

- it covers *language*.
- it involves a lot of *formal* (i.e. mathematical) modeling—language is inherently formal!

This course differs from other *linguistics* courses because:

- it focuses on linguistic 'performance' rather than linguistic 'competence' [Chomsky, 1965].
  - **competence**: mental representations of linguistic knowledge (rules to combine signs)
  - **performance**: how language is actually used (regularities in how speech errors happen)
- it models phenomena at an 'algorithmic' rather than 'computational' level [Marr, 1982].
  - **computational/functional**: model the task a behavior does, e.g. find spoken phrases.
  - **algorithmic/representational**: model processes/structures behaviors use, e.g. memory.
  - **implementational**: model physical implementation of behaviors, e.g. neural firing.

The course therefore covers some of the same material as other linguistic courses, but differently.

The course is organized into three parts:

1. background (what we will assume about how the brain works):
   - neural firing, mental states, cued associations, complex ideas
2. the processes of language:
   - **decoding** complex signs into complex ideas
     - identifying words and phrases and associating them with meanings
   - **encoding** complex ideas into complex signs
     - turning meanings back into words and phrases

3. acquisition (how babies learn language):
    - learning speech sounds
    - learning words and meanings
    - learning to encode and decode complex ideas

1.2 Background: some math notation (in case you don't know)

Set notation, involving **sets** S, S' and **entities** $x, x', x'', x_1, x_2, x_3, \ldots$:

pair	$\langle x_1, x_2 \rangle$
tuple	$\langle x_1, x_2, x_3, ... \rangle$
set	$S = \{x \mid ...\}$ e.g. $\{x_1, x_2, x_3\}$
empty/null set	$\emptyset$ or $\{\}$
element	$x \in S$ e.g. $x_2 \in \{x_1, x_2\}$, $x_3 \notin \{x_1, x_2\}$
subset (or equal)	$S \subset S'$ e.g. $\{x_1, x_2\} \subset \{x_1, x_2, x_3\}$, $\{x_1, x_2\} \subseteq \{x_1, x_2\}$
union	$S \cup S'$ e.g. $\{x_1, x_2\} \cup \{x_2, x_3\} = \{x_1, x_2, x_3\}$
intersection	$S \cap S'$ e.g. $\{x_1, x_2\} \cap \{x_2, x_3\} = \{x_2\}$
exclusion or complementation	$S - S'$ e.g. $\{x_1, x_2\} - \{x_2, x_3\} = \{x_1\}$
Cartesian product	$S \times S'$ e.g. $\{x_1, x_2\} \times \{x_3, x_4\} = \{\langle x_1, x_3\rangle, \langle x_1, x_4\rangle, \langle x_2, x_3\rangle, \langle x_2, x_4\rangle\}$
power set	$\mathcal{P}(S)$ or 2^S e.g. $\mathcal{P}(\{x_1, x_2\}) = \{\emptyset, \{x_1\}, \{x_2\}, \{x_1, x_2\}\}$
relation	$R \subseteq S \times S' = \{\langle x, x'\rangle \mid ...\}$ e.g. $R = \{\langle x_1, x_3\rangle, \langle x_2, x_3\rangle, \langle x_2, x_4\rangle\}$
function	$F : S \rightarrow S' \subseteq S \times S'$ s.t. if $\langle x, x'\rangle, \langle x, x''\rangle \in F$ then $x' = x''$
cardinality	$\lvert S \rvert$ = number of elements in S
real numbers	$\mathbb{R}$: the uncountably infinite set of real numbers
real ranges	$\mathbb{R}_m^n$: the real numbers between m and n (inclusive)
real tuples	$\mathbb{R}^n$: the uncountably infinite set of n-tuples of reals

First-order logic notation, involving **propositions** p, p' – e.g. that $1<2$ (true) or $1=2$ (false):

conjunction	$p \wedge p'$ or p, p' e.g. $1<2 \wedge 2<3$ or $1<2, 2<3$: both p and p'
disjunction	$p \vee p'$ e.g. $1<2 \vee 1>2$: either p or p'
negation	$\neg p$ or '/' e.g. $\neg 1=2$ or $1 \neq 2$: not p
implication	$p \rightarrow p'$ (equivalent to $\neg p \vee p'$) e.g. 3 is prime $\rightarrow$ 3 is odd
existential quantifier	$\exists_{x \in S} \ldots x \ldots$: disjunction over all x of proposition $\ldots x \ldots$
universal quantifier	$\forall_{x \in S} \ldots x \ldots$: conjunction over all x of proposition $\ldots x \ldots$

Limit notation, involving **sets** S and **entities** x:

limit union	$\bigcup_{x \in S} \ldots x \ldots$:	union over all x of set $\ldots x \ldots$
limit intersection	$\bigcap_{x \in S} \ldots x \ldots$:	intersection over all x of set $\ldots x \ldots$
limit sum	$\sum_{x \in S} \ldots x \ldots$:	sum over all x of number $\ldots x \ldots$
limit product	$\prod_{x \in S} \ldots x \ldots$:	product over all x of number $\ldots x \ldots$
limit	$\lim_{x \rightarrow \infty} \ldots x \ldots$:	limit as x tends to infinity of number $\ldots x \ldots$

1.3 Background: probability and probability spaces [Kolmogorov, 1933]

Probability is defined over a measure space $\langle O, \mathcal{E}, \mathsf{P}\rangle$ where the measure P (probability) sums to one.

This **probability measure space** $\langle O, \mathcal{E}, \mathsf{P}\rangle$ consists of:

1. a **sample space** O – a non-empty set of **outcomes** (e.g. the numbers on a die);
2. an **event space** ('sigma-algebra') $\mathcal{E} \subseteq 2^O$ – a set of **events** in the power set of O such that:
   (a) $\mathcal{E}$ contains O: $O \in \mathcal{E}$ (e.g. the event of rolling *any* number: $\{1, 2, 3, 4, 5, 6\}$ is in $\mathcal{E}$),
   (b) $\mathcal{E}$ is closed under complementation: $\forall_{A \in \mathcal{E}}\ O{-}A \in \mathcal{E}$ (e.g. rolling no number: $\emptyset$ is in $\mathcal{E}$),
   (c) $\mathcal{E}$ is closed under countable union: $\forall_{A_1..A_\infty \in \mathcal{E}} \bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$ (if $\{1, 2\}$ and $\{3\}$ then $\{1, 2, 3\}$)

   (this set of events will be the **domain** of our probability function – things with probability);
3. a **probability measure** $\mathsf{P} : \mathcal{E} \to \mathbb{R}_0^\infty$ – a function from events to non-negative reals such that:
   (a) the P measure is countably additive: $\forall_{A_1..A_\infty \in \mathcal{E} \text{ s.t. } \forall_{i,j}\ A_i \cap A_j = \emptyset}\ \mathsf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$,
   (b) the P measure of entire space is one: $\mathsf{P}(O) = 1$ (e.g. P(rolling any number) = 1).

This characterization is helpful because it unifies probability spaces that may seem very different:

1. **discrete** spaces – e.g. a coin:

   $$\langle \underbrace{\{\mathrm{H}, \mathrm{T}\}}_{O}, \underbrace{\{\emptyset, \{\mathrm{H}\}, \{\mathrm{T}\}, \{\mathrm{H}, \mathrm{T}\}\}}_{\mathcal{E}}, \underbrace{\{\langle \emptyset, 0\rangle, \langle \{\mathrm{H}\}, .5\rangle, \langle \{\mathrm{T}\}, .5\rangle, \langle \{\mathrm{H}, \mathrm{T}\}, 1\rangle\}}_{\mathsf{P}}\rangle$$

2. **continuous** spaces – e.g. a dart (here $2^{\mathbb{R}^2}$ is a Borel algebra: a set of all open subsets of $\mathbb{R}^2$):

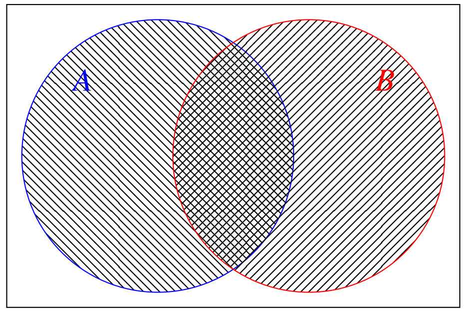
   $$\langle \underbrace{\mathbb{R}^2}_{O}, \underbrace{2^{\mathbb{R}^2}}_{\mathcal{E}}, \underbrace{\{\langle R, p\rangle \mid R \in 2^{\mathbb{R}^2}, p = \iint_{A \in R} \mathcal{N}_{\mathbf{0},\mathbf{1}}(x_A, y_A)\, dA\}}_{\mathsf{P}}\rangle$$

   (events must be open sets/ranges of outcomes because point outcomes have zero probability)
3. **joint** spaces using Cartesian products of sample spaces – e.g. two coins ($\{\mathrm{H}, \mathrm{T}\} \times \{\mathrm{H}, \mathrm{T}\}$):

   $$\langle \underbrace{\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}}_{O}, \underbrace{\{\emptyset, \{\mathrm{HH}\}, \ldots, \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}\}}_{\mathcal{E}}, \underbrace{\{\langle \emptyset, 0\rangle, \langle \{\mathrm{HH}\}, .25\rangle, \ldots, \langle \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}, 1\rangle\}}_{\mathsf{P}}\rangle$$

This axiomatization entails, for any events $A, B \in \mathcal{E}$ (e.g. rolling an even number or less than 4):

1. $\mathsf{P}(A) \in \mathbb{R}_0^1$
2. $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$



Though probabilities are defined over sets of outcomes, we often write them using **propositions**.

For example, if $O = X \times Y$ (say, flipping a coin and rolling a die) and therefore $\forall_{o \in O}\ o = \langle x_o, y_o \rangle$:

$$\begin{aligned}
\mathsf{P}(x) &= \mathsf{P}(X{=}x) &&= \mathsf{P}(\{o \mid o{\in}O \land x_o{=}x\}) && \text{(allow any value for } y_o \text{ component)}\\
\mathsf{P}(x \land y) &= \mathsf{P}(X{=}x \land Y{=}y) &&= \mathsf{P}(\{o \mid o{\in}O \land x_o{=}x \land y_o{=}y\})\\
\mathsf{P}(\neg x) &= \mathsf{P}(X{\neq}x) &&= \mathsf{P}(\{o \mid o{\in}O \land x_o{\neq}x\})
\end{aligned}$$

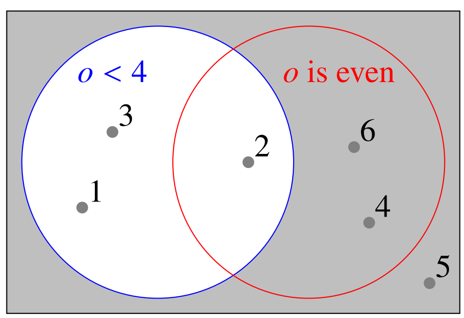
Random variables D are functions from outcomes x_o, y_o to **values** (e.g. distance of point to origin).

Often we simply use the Cartesian factors of a joint sample space (X, Y) as random variables.

We can also define **conditional probabilities** as ratios of these measures: $\mathsf{P}(S \mid R) = \frac{\mathsf{P}(R \cap S)}{\mathsf{P}(R)}$.

(It's the probability of the joint or intersection $R \cap S$ over the probability of the condition R.)

For example, if we have a fair die roll, then $\mathsf{P}(o \text{ is even} \mid o < 4) = \frac{\mathsf{P}(o \text{ is even } \land\ o<4)}{\mathsf{P}(o<4)} = \frac{1}{3}$:



Practice 1.1:

Using variables X and Y for two coin flips, each with outcomes H and T, write a probability equation expressing that a quarter of the time the first coin will come up heads and the second coin will come up tails.

Practice 1.2:

Assuming two fair coins are tossed, each with a .5 probability of a heads outcome and a .5 probability of a tails outcome, what is the probability that at least one coin will come up heads?

References

[Chomsky, 1965] Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

[Kolmogorov, 1933] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. Second English Edition, *Foundations of Probability* 1950, published by Chelsea, New York.

[Marr, 1982] Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Company.