

LING3701/PSYCH3371: Lecture Notes 1

Introduction and Background

Contents

1.1	What is this course about?	1
1.2	Why study psycholinguistics?	2
1.3	Background: some math notation (in case you don't know)	2
1.4	Background: probability and probability spaces [Kolmogorov, 1933]	3
1.5	Practice: notation	4
1.6	Practice: probability calculation	4

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 Why study psycholinguistics?

It's interesting: Conscious thought is still poorly understood. Language is a tool to observe it.

It's useful: You use language every day to communicate. It may help to know what can go wrong.

It's a living: You may end up wanting to do research in this field. It happened to me...

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

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

pair	$\langle x_1, x_2 \rangle$
tuple	$\langle x_1, x_2, x_3, \dots \rangle$
set	$S = \{x \mid \dots\}$ 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	$R \subseteq S \times S' = \{\langle x, x' \rangle \mid \dots\}$ 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	$ S $ = 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 $1 < 2, 2 < 3$
disjunction	$p \vee p'$ e.g. $1 < 2 \vee 1 > 2$
negation	$\neg p$ or '/' e.g. $\neg 1 = 2$ or $1 \neq 2$

implication	$p \rightarrow p'$ (equivalent to $\neg p \vee p'$) e.g. 3 is prime \rightarrow 3 is odd
existential quantifier	$\exists_{x \in S} \dots x \dots$: disjunction over all x of $\dots x \dots$
universal quantifier	$\forall_{x \in S} \dots x \dots$: conjunction over all x of $\dots x \dots$

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

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

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

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

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

1. a **sample space** O – a non-empty set of **outcomes**;
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 : $O \in \mathcal{E}$,
 - (b) \mathcal{E} is closed under complementation: $\forall A \in \mathcal{E} \ O - A \in \mathcal{E}$,
 - (c) \mathcal{E} is closed under countable union: $\forall A_1..A_\infty \in \mathcal{E} \ \bigcup_{i=1}^\infty A_i \in \mathcal{E}$

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

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

1. **discrete** spaces – e.g. a coin:

$$\underbrace{\langle \{H, T\} \rangle}_O, \underbrace{\langle \emptyset, \{H\}, \{T\}, \{H, T\} \rangle}_\mathcal{E}, \underbrace{\langle \langle \emptyset, 0 \rangle, \langle \{H\}, .5 \rangle, \langle \{T\}, .5 \rangle, \langle \{H, T\}, 1 \rangle \rangle}_P$$

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):

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

(events must be open sets/ranges of outcomes because point outcomes have zero probability)

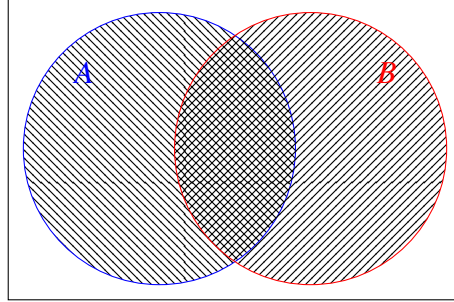
3. **joint** spaces using Cartesian products of sample spaces – e.g. two coins ($\{H, T\} \times \{H, T\}$):

$$\underbrace{\langle \{HH, HT, TH, TT\} \rangle}_O, \underbrace{\langle \emptyset, \{HH\}, \dots, \{HH, HT, TH, TT\} \rangle}_E, \underbrace{\langle \langle \emptyset, 0 \rangle, \langle \{HH\}, .25 \rangle, \dots, \langle \{HH, HT, TH, TT\}, 1 \rangle \rangle}_P$$

(sample spaces in Cartesian products of joint probabilities are called **random variables**)

This axiomatization entails, for any $A, B \in \mathcal{E}$:

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



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

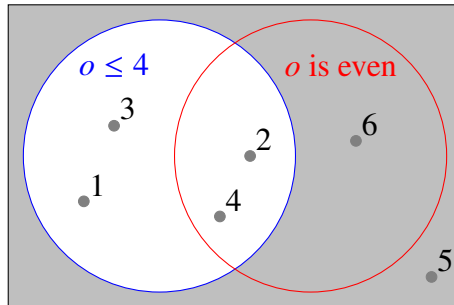
For example, if $O = X \times Y$ and therefore $\forall_{o \in O} o = \langle x_o, y_o \rangle$:

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

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

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

For example, if we have $O = \{1, 2, 3, 4, 5, 6\}$, then $P(o \text{ is even} \mid o \leq 4) = \frac{P(o \text{ is even} \wedge o \leq 4)}{P(o \leq 4)} = \frac{2}{4} = \frac{1}{2}$.



1.5 Practice: notation

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.

1.6 Practice: probability calculation

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*. MIT Press, Cambridge, Mass.
- [Kolmogorov, 1933] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin. 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.