Introduction to experimental methods for linguists

Nausicaa Pouscoulous
University College London

Outline of the workshop

Lecture 1:
Why should linguists do experiments?
How to read an experimental paper in linguistics?

Lecture 2:
Introduction to experimental methods
Experimental design for linguists

Lecture 3: Types of experimental procedures

Why should linguists do experiments?

with empirical methods: get objective data - not biased by intuitions of linguistically trained users

experiments allow study of linguistic processing below the level accessible to introspection:
e.g., time-course of inferential processing which reflects the stages of processing rather than the eventual interpretation of a certain expression.
Why should linguists do experiments?

in turn, both can help us evaluate (and discriminate between) linguistic theories (especially those with a psychological aspect) and where reflective intuition alone is not sufficient.

empirical data can also lead us to modify a theory (top-down vs. bottom-up research)

Acceptability judgments

Check theorists’ intuitions about acceptability of sentences
Acceptability, grammaticality, naturalness, comprehensibility, felicity, appropriateness…
Aren’t theorists’ intuitions solid?

Wrong intuitions
Scalar implicatures

Examples:
some < most < all
or < and
warm < hot
may < must

In many cases, the use of a weaker scalar term implies the negation of the stronger ones:
Anna ate some of the cookies => She didn’t eat all of them.
You can eat cheese or dessert => You can’t have both.
It’s warm outside => It isn’t hot outside.
You may kiss the bride => You don’t have to.

Wrong intuitions
Scalar implicatures

Scalar inferences arise when a speaker chooses to use a relatively weak term rather than a stronger, more informative one from the same semantic scale. The hearer thereby understands that the stronger term could not be used.

There are two possible interpretations of weak scalar terms (such as or and some):
without the scalar inference they are compatible with the stronger term (and, all) => plain meaning (logical interpretation)
with the scalar inference, they deny the stronger term => scalar reading (pragmatic interpretation)

Wrong intuitions
Scalar implicatures

Plain meaning (literal meaning): scalar terms (or, some) are compatible with the stronger term (and, all).
Pragmatic interpretation (with the scalar inference): they deny the stronger term (not both, not all).

Wrong intuitions
Scalar implicatures

Pouscoulous et al. (2007) - Method Experiment 2

D1 : I would like all the boxes to have a token.
D2 : I would like some (quelques) boxes to have a token.
D3 : I would like none of the boxes to have a token.
D4 : I would like some of the boxes not to have a token.

Situation 1

Situation 2

Situation 3
**Wrong intuitions**  
Pouscoulos et al. (2007) - Results Experiment 3

Quelques vs. Certains

I would like some of the boxes to have a token.

Key scenario

---

Adults derive implicature with both quantifiers: quelles (83%), certains (79%)  
9-10-year-olds derive implicature more with quelles (100%) than with certains (52%)

---

**Invisible psychological processes & deciding between theories**

Are scalar implicature derived by default?  
Are they derived by default and canceled if necessary?  
Or are they inferred on-line only when the context licences them?  
Same outcome in both types of theories, but different time-course predicted

---

**Focus questions**  
Kastos & Cummins (2010)

1. How important is it that a linguistic theory should be psychologically plausible?  
2. How can we prove the psychological reality of theoretical constructs, such as implicature, presupposition etc.? Can we ever disprove their psychological reality?  
3. Do we elicit naturalistic linguistic behaviour under experimental conditions? In what ways might it be unrealistic or unrepresentative?  
4. How can we operationalise the predictions of psychological theories? For example, how do we experimentally measure ‘effort’ in reasoning? What might confound these measurements?  
5. What are the advantages and disadvantages of gathering syntactic/semantic/pragmatic intuitions from a broad array of participants rather than linguistically-trained specialists?  
6. How should we interpret variability between participants in their responses or intuitions?

---

**How to read an empirical paper**
How to read and empirical paper
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

Based on a guide by Wing-Yee Chow and Shevaun Lewis
Empirical papers generally organized into: Introduction, Methods, Results, and Discussion. (Sometimes Methods at the end of the paper, but better to read the Methods before the Results). Each section contains very different kinds of information and it is most effective to read each of them with a clear set of goals in mind.
You should try to answer these questions when you are reading:

How to read the Introduction
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

What are the big-picture questions?
A good paper starts by explaining what is at stake first. However, at times the big-picture questions are not immediately clear in the introduction. If the authors did not state the big-picture questions explicitly, you should try to come back to this after reading the entire Introduction section. Sometimes the authors talk about a debate with opposing points of view, and the question is implicit.

How to read the Introduction
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

What specific question(s) did the authors try to ask in this paper?
Authors of empirical papers approach the big picture questions by formulating more specific questions that can be addressed experimentally. You should identify the authors’ arguments for why this is an interesting/reasonable question to ask.

How to read the Introduction
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

How does the current study contribute something new compared to what previous studies have accomplished before?
Most Introduction sections provide a review of the previous work on related research questions. You should try to identify how the current study can potentially add to our knowledge about such questions.

How to read the Introduction
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

What hypotheses did the authors discuss?
Hypotheses are about underlying mental processes (e.g., processes involved in sentence comprehension) and are independent of the experimental methods. Important: Do NOT confuse hypotheses with predictions. Predictions describe the kind of evidence one would observe within a particular experiment if the hypotheses were correct.

Some possible scenarios:
Completely new research question - no one has tried to answer it before (pretty unusual).
Clarify previous findings: previous findings are missing pieces, the results are in conflict, are compatible with multiple explanations, etc., → the current experiment intends to clarify what the previous findings mean.
Extend previous findings to a new population/environment (e.g., speakers of different languages, children, second language learners).
What is the experimental design?
Identify the independent and dependent variables

Experiments involve manipulating one or more factors and measuring the effects of such manipulations on observable outcomes. The factors that are manipulated are called “independent variables”, and the outcomes that are measured are “dependent variables”. Each factor has multiple levels, often called conditions.

Example: We want to know the effects of temperature on a person’s heart rate. We can manipulate the temperature of the environment (e.g., high vs. low temperature). In this case, temperature of the environment is an independent variable / factor, and it has two levels (high temperature condition and low temperature condition). We can then measure people’s heart rates (a dependent variable) in each condition.

Try to draw a table to explain to yourself the experimental design and fill in the boxes with sample items.

Describe what a trial would be like in each condition (i.e., what a participant would see/hear in each condition).

Example:

<table>
<thead>
<tr>
<th>Plurality</th>
<th>Animacy</th>
<th>Animate</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>Cat</td>
<td>Car</td>
<td></td>
</tr>
<tr>
<td>Plural</td>
<td>Cats</td>
<td>Cars</td>
<td></td>
</tr>
</tbody>
</table>

Sampling from a population

How many participants were tested?
What are some characteristics of the population of interest?
Were they college students/ patients/ children / infants?

How many items per condition participants see/hear?
Either reported directly or you need to figure it out (from total number of items and number of conditions).

Big differences depending on tested population(e.g., healthy adults can sit through a 2-hour session, not infants/toddlers) and the experimental method used (e.g., ERP studies tend to require a much larger number of data points than self-paced reading).

What are the characteristics of the materials?
The stimuli across conditions will inevitably be different in many aspects besides those manipulated as the independent variable. For example, when you manipulate the plurality of a noun, e.g., car – cars, you are inevitably also changing the length/spelling of the word.

What unintended variation is controlled for? How was it controlled for, i.e., by being held constant or by counterbalancing?
How to read the Methods
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

What was the task?
Some tasks are more common/simple than others, but at times you need to first understand the task in the experiment before being able to identify where the dependent variables come from (e.g., reaction time to what? how was it measured?)
You can usually get the most detailed description of the task in the subsection called “Procedure”, but note that it usually comes after the description of the experimental design and materials.

How to read the Methods
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

What does each of the hypotheses predict?
If an experiment is well-designed, the competing hypotheses should make contrasting predictions. You should try to figure out the predictions yourself, and then compare them to what the authors said the predictions are. Sometimes the authors put this in the Introduction section.
* Understanding the predictions of the hypotheses is one of the most important steps in understanding an experiment.

How to read the Results
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

Understand the figures in the paper and link them back to the predictions.
Results sections are often filled with technical details about the results of the statistical analyses. We’ll talk about some (e.g., p-values), but don’t get blocked on the details of the statistics.
Rather, it is important that you try to understand the descriptions of the result pattern and be able to read the figures to identify whether the results match up with the predictions.

How to read the Discussion
Wing-Yee Chow and Shevaun Lewis (Fall 2011)

Before reading the Discussion, think about what the results mean and how they bear on the research question.
The author’s interpretation of the results
First a summary of the findings.
Then authors free to discuss what the findings mean --> decide for yourself whether you agree with their interpretation.
The authors sometimes discuss questions that arise from their findings, which can provide interesting insights for future research.
**Critical reading for non-experts**

- Style and general readability of the paper – is it easy to understand?
- Is the theory accurately portrayed?
- Do the experimental prediction correspond to the theoretical hypothesis? (Translation btw theory and experiments)
- Is the experimental design good? Is it suitable for the research question put forward? (Does it test the hypothesis?)

---

**Critical reading for non-experts**

- Is the review of the experimental literature good? Are the other important studies in the domain mentioned when relevant?
- Are the conclusions the authors draw legitimate in view of the data?
- Are there other possible interpretations?

---

**Critical reading for non-experts**

- Missing stats
- Missing drop outs
- Exclusion criteria
- Counterbalancing
- Interobserver reliability, if relevant

---

**Case studies**

**Paul Grice’s Cooperative Principle**

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

(Grice 1975/1989: 26)

---

**Case studies**

**Paul Grice’s Conversational Maxims**

Maxims of quantity
1. Make your contribution as informative as is required (for the current purpose of the exchange).
2. Do not make your contribution more informative than is required.

Maxims of quality
Supermaxim: Try to make your contribution one that is true
1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Maxim of relation
Be relevant.

Maxims of manner
Supermaxims: Be perspicuous
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.
Case studies
Grice’s inferential procedure to derive implicatures

The speaker utters p.

H. has no reason to assume that S doesn’t observe the conversational maxims or at least the Cooperative Principle.

The assumption that the S obeys the CP and the maxims require the assumption that he thinks that q.

• The S knows (and knows that the H knows that he knows) that the S understands that it is necessary to assume that the S thinks that q.
• The S did nothing to prevent the H to think that q.
• The S wants the H to think that q.
• The S has implicated q.

Case studies
Gricean metaphor processing model

Literal meaning should be arrived first - metaphors subsequent (and slower)
processing the metaphorical interpretation should be optional

Experimental research shows both claims wrong
the understanding of metaphorical expressions in context can be as easy and fast as that of a literal turn of phrase
it is not optional, but mandatory and automatic

Case studies
Grice’s analysis of metaphor

Metaphors are violations of maxim of truthfulness, designed to convey a related true implicature.

The three-stage model of metaphor comprehension predicts that:

hearing will expect literal truthfulness (as required by the maxim of truthfulness) and they should therefore try a literal interpretation first.
If this reading is subsequently found to be nonsensical in context,
this will, in turn, prompt them to look for a contextually relevant non-literal alternative

Case studies
Link between theory and predictions

Grice wanted reconstruct a plausible derivation of implicatures, not to provide an explanatory account of communication from a psychological point of view.

See, Geurts & Rubio-Fernández (2015)

Case studies
David Marr (1982)

Difference between three different levels of analysis: computational, algorithmic/representational and implementational

computational level: what? Which are the input and output of the process + the constraints that allow a specified problem to be solved.

representational level: How? How we get from input to output + which representations have to be used and which processes have to be employed in order to build and manipulate the representations.

implementational level: description of the physical system that realises the process at the physical e.g., neuronal level.

Case studies
Scalar implicatures development

Noveck (2001) - Experiment 2

He tested adults and children (8 and 10 years-old) using sentences which would be considered true if ‘some’ was understood to be compatible with ‘all’, but false if the inference was drawn, such as:

‘some elephants have trunks’
‘some giraffes have long necks’
Children generally did not make the scalar inference whereas adults were equivocal.

<table>
<thead>
<tr>
<th>% pragmatic responses</th>
<th>7-8-year-olds</th>
<th>10-11-year-olds</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some giraffes have long neck</td>
<td>11</td>
<td>15</td>
<td>69</td>
</tr>
</tbody>
</table>

Case studies
Scalar implicatures development
Noveck (2001) - Experiment 2

Do 5-year-olds (4;1 to 6;1; M age = 5;3) understand scalar implicatures?
Rewarding game: puppet gets reward if it completes the activity.
3 conditions:
- Quantificational condition (n=10):
  E: Did you color the stars? Elephant: I colored some. 
  E: Did you eat the sandwich? Bear: I ate the cheese. 
- Encyclopaedic condition (n=10):
  E: Did you wrap the gifts? Cow: I wrapped the parrot. 
  No reward in 77.5%, 70% and 90% (respectively).

Case studies
Papafragou & Tantalou (2004)
A problem?
Do 5-year-olds (4;1 to 6;1; M age = 5;3) understand scalar implicatures?
Rewarding game: puppet gets reward if it completes the activity.
3 conditions:
- Quantificational condition (n=10):
  E: Did you color the stars? Elephant: I colored some. 
  E: Did you eat the sandwich? Bear: I ate the cheese. 
- Encyclopaedic condition (n=10):
  E: Did you wrap the gifts? Cow: I wrapped the parrot. 
  No reward in 77.5%, 70% and 90% (respectively).

References
Berger, F. & Höhle, B. (2012). Restrictions on addition. Children's interpretation of the focus particles auch (also) and nur (only) in German. Journal of Child Language. 39, 383-410.
Chow W.Y., & Lewis, S. (Fall 2011). How to read an empirical paper. For their LING499E Real-time Language Processing course.