Grammaticality Judgments as Linguistic Evidence
Alternatives to Judgements

Brian Murphy

Language, Interaction and Computation Lab
Centre for Mind/Brain Sciences (CIMeC)
University of Trento

ESSLLI 2009, Bordeaux
Course Outline

- Notions of Grammaticality, and Current Practice in Linguistics

**Others Sources of Linguistic Evidence**
- Behavioural experiments: timing, eye-tracking
- Neuroscience: EEG, MEG, fMRI
- Usage and frequency

- Methodology for Eliciting Judgements of Grammaticality
- Scales for Measuring Grammaticality
- Theoretical Implications
Linguistics aims to describe the grammars that result in the structures and other patterns we see in particular languages

- Focus on rules and representations
- Uses people to find out about language
Psycholinguistics investigates the processes that allow us to produce and comprehend language, and more broadly to understand cognitive functions (e.g. attention, learning, memory) that are shared with other human capabilities.

- Focus on processes and timing
- Uses language to find out about people
As a result, the tools psycholinguists use are often not the best choice for linguists.

Here we’ll concentrate on studies of sentence comprehension, since this is where you can see “violations”
Caveat: many (psycho)linguists do very interesting cross-over work on the architecture of language, that uses a combination of methodological approaches and is of direct relevance to both communities:

- visual world paradigm,
  [Tanenhaus et al., 1995, Karabanov et al., 2007]: interpretation proceeds before the end of a phrase is reached, including integration of information from the (visual) context, and pronoun reference binding

- violation ERP paradigm,
  [Hagoort et al., 2004, Özyürek et al., 2007, Nieuwland et al., 2007]: semantic congruence violations can be triggered by “extralinguistic” encyclopaedic knowledge, by meaningful gestures, and depend on current situational information
Timing-Based Paradigms I

- Timed grammaticality judgements
  - not a simple relationship: clearly grammatical or ungrammatical sentences are fast; marginal or ambiguous sentences are slow
  - little used, outside acquisition research [McDonald, 2000, Kail, 2004]
  - [Bader and Häussler, ] note that relationship can hold among separately in grammatical and ungrammatical sets of sentences

Figure A1
Mean SGJ reaction times for judgments ‘grammatical’ (filled circles) and judgments ‘ungrammatical’ (open circles) plotted against perceived well-formedness. Reaction times are from Experiment 2. Left: Mean acceptability as obtained in the ME part of Experiment 1. Right: Mean grammaticality as obtained in the SGJ part of Experiment 2.
Self-paced reading/listening

- sentences that require more processing, will take longer to read - due to structural complexity, or ambiguity
- informants proceed through sentence, word by word
- allows us to see in where in the sentence processing load is higher
- again, is orthogonal to grammaticality
Eye-tracking

- same rationale as self-paced reading [Rayner, 1998]
- can see number and duration of gaze-fixation on individual words
- additionally, allows us to observe when alternative analyses are entertained
- reading more naturalistic than in self-paced experiments
- again, effects may or may not be related to grammaticality
Timing-Based Paradigms IV
When neurons fire, they produce tiny electrical currents
Much of the time, these currents are not in sync, and largely cancel each other out.

Also subject to many non-brain artefacts, though there are methods for dealing with them.
Neuroscience: EEG III
Neuroscience: EEG IV

- **Event-Related Synchronisation:**
  - certain cognitive functions cause networks of cells to synchronise (gamma, >20Hz)
  - lack of function can cause neuronal idling (alpha, ~10Hz)
  - but no established links to grammaticality (though very useful for all sorts of other things)

- **Event-Related Potentials**
  - certain functions cause a phase-reset, time-locked to the stimulus, which emerges with grand averages over sentences and participants
  - certain Event-Related Potentials (ERPs), are associated with ungrammaticality

  ★ see [Osterhout and Nicol, 1999] for “violation” ERPs
  ★ and [Kutas et al., 2006] for EEG in psycholinguistics generally

http://www.u.arizona.edu/~vanpettc/reprints/Psycholinguistics_Electrification
Resulting voltages at the scalp are a dulled version of what is happening within the brain.

Analogies:
- hearing a party in a neighboring apartment
- hearing a crowd at a football stadium

Excellent timing resolution (ms), coarse spatial resolution (cm)
EEG: Syntax, P600

- Syntactic violations
- Garden-path or other complex and processing-intensive sentences
- Detection, reanalysis, repair of a violation

The cat will EAT

*The cat will EATING

Osterhout & Nicol 1999
Implausibility, semantic incongruence, e.g. breaking selectional restrictions [Kutas and Hillyard, 1980, Osterhout and Nicol, 1999]
EEG: Semantics, Pragmatics, N400 II

- Size of deflection related to improbability
  [Kutas and Federmeier, 2000]
A thematic incongruity can trigger a reanalysis, resulting in a “syntactic” P600 [Kim and Osterhout, 2005]
EEG: Complications II

- Unexpected syntax in argument structure can trigger a “semantic” N400 [Bornkessel and Schlesewsky, 2006]
EEG: Practicalities

- Preparation can take 30mins to an hour
- Recordings often done in an electrically shielded cabin, which some may find claustrophobic
- Participants are asked not to move, particularly face/head (swallowing, eye-movements, grinding teeth, etc)
- Maximum reasonable experiment duration ca. 60mins
- Typically need 20-50 sentences per condition, and 10-20 participants
- Total cost to organisation: for us ca. 30 euros per hour = 60 euros per session = 600-1200 per study (though considerably cheaper systems are available, and may soon be available as consumer products)
MEG

- Measures much the same activity as EEG, but via corresponding magnetic field produced by brain
- More sensitive; spatial resolution much better; same temporal resolution as EEG; higher frequencies; not affected by skull and scalp; easier to compare individuals; more comfortable for participants (sitting position, can move)
- Much more expensive (uses superconducting sensors > liquid helium cooling)
fMRI

- Active method - participant is in very strong standing magnetic field (superconducting), and radio waves are used to measure physical properties of brain.
- Much better spatial resolution (mm); sees through skull; easy to generalise across participants.
- Much worse temporal resolution (5-10s), very very expensive (our lab, ca. 300 euros per hour), very loud and somewhat claustrophobic.
- Some work on grammaticality - findings that different brain locations respond to sentential complexity, to ungrammaticality, and to semantic anomalies [Ni et al., 2000, Newman et al., 2001, Friederici et al., 2006]
Usage and Frequency I

- Relationship to grammaticality not well understood
- ... though some have posited that (frequency derived) probability of a structure, or string of words, may drive grammaticality [Manning and Schütze, 1999]
- Cloze probability corresponds to grammaticality in selectional restrictions (above)
- Perhaps more suited to quantifying optionality?
- Is optionality orthogonal to grammaticality, or are they on the same continuum?
- Ceiling and floor effects may be responsible for confusion
- If there is a ‘sweet point’ between ceiling and floor effects, how do you find it?
Usage and Frequency II

- [Bader and Häussler, ]: active, passive and recipient-passive with a range of verbs

  - ... dass dem Bub (vom Onkel) ein Ring geschenkt wurde
    ... that the.DAT boy (by-the uncle) a.NOM ring presented was
    ‘... that a ring was presented to the boy (by the uncle)’

  - ... dass der Bub (vom Onkel) einen Ring geschenkt bekam
    ... that the.NOM boy (by-the uncle) a.ACC ring presented got
    ‘... that the boy was presented a ring (by the uncle)’

  - ? Ich weiß, dass dem Mann ein Buch gestohlen bekam
    I know that the man a book stolen was
    ‘I know that the man was stolen a book’

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Regular Passive</th>
<th>Recipient Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ME × log(V+Aux)</td>
<td>.06</td>
<td>.08</td>
<td>.52**</td>
</tr>
<tr>
<td>(2) SGJ × log(V+Aux)</td>
<td>.19</td>
<td>.20</td>
<td>.34*</td>
</tr>
<tr>
<td>(3) ME × log(V+Aux/V)</td>
<td>-.22</td>
<td>.16</td>
<td>.49**</td>
</tr>
<tr>
<td>(4) SGJ × log(V+Aux/V)</td>
<td>.00</td>
<td>.03</td>
<td>.48**</td>
</tr>
</tbody>
</table>
Usage and Frequency III

- Compare Subject-Object ordering in the (normal \textit{wurden}) passive, varying object case and animacy

- Compare sentence final verb cluster ordering
  
  Ich glaube, dass Karl das Buch \textbf{hat kaufen müssen}
  
  I believe that K. the book has buy must
  
  ‘I believe that Karl had to buy the book’

\textbf{Figure 8}

Comparison of corpus data with experimental results (ME results of Experiment 1 and offline grammaticality judgments (OGJ) of Experiment 3).

Percentage scale on left y-axis and ME scale on right y-axis.

Left graph: Order between subject and object (Subexperiments A and B combined).

'Sv/O-: Ace' is short for 'subject[:animate] before object[:animate, accusative]' and similar for the other conditions. Right graph: Verb-cluster serialization (Subexperiment C)
Psycholinguists and Linguists are often interested in different aspects of the same problem.

So their methods are not always appropriate for each other.

Certainly, grammaticality (judgement) experiments are vulnerable to participant bias, in a way that many other paradigms are not.

But in using more objective methods, additional assumptions are required during interpretation (e.g., what does the N400 mean? With judgements you can ask/instruct people).
Current psycholinguistic paradigms are best suited to questions about the architecture of the language faculty and are not sensitive enough to answer specific questions about grammar (at least not practicably).

- Don’t worry (about using intuitions) ...
- ... be happy (that it is the best one for your purposes)
- Discussion ...
What Comes Next

- Notions of Grammaticality, and Current Practice in Linguistics
- Others Sources of Linguistic Evidence
- Methodology for Eliciting Judgements of Grammaticality
- Scales for Measuring Grammaticality
- Theoretical Implications
Acknowledgments

- Francesco Vespignani, CImeC, from whose slides this presentation borrows
- Fellow group members Roberto Zamparelli, Marco Baroni, Massimo Poesio, and all participants in Linguistic Evidence seminar
References I

Bader, M. and Häussler, J. 
Toward a model of grammaticality judgments. 
under review.

Generalised semantic roles and syntactic templates: A new framework for language comprehension. 

Processing linguistic complexity and grammaticality in the left frontal cortex. 
*Cerebral Cortex*, 16(12):1709–1717.

Integration of word meaning and world knowledge in language comprehension. 

On-line grammaticality judgments in French children. 
References II


References III


