Magnitude Estimation for Linguistic Acceptability

Dora Alexopoulou
Marie Curie Fellow
CNRS Lab Savoirs, Textes, Langage
Lille III
Overview

- The problem
- Magnitude Estimation for linguistic acceptability
- Survey of recent applications
- WebExp

The problem

(i) Who does John think Mary will fire t?
(ii) ?*Who did Mary wonder whether they will fire t?
(iii) *Who did John meet the girl who will marry t?

What is the status of sentences of intermediate acceptability like (ii)?
(i) Who does John think Mary will fire t?

(ii) *Who did Mary wonder whether they will fire t?

(iii) *Who did John meet the girl who will marry t?

(v) Who were you wondering if we should see?
(Chung &McCloskey 1983).
Are such discrepancies an artefact of the absence of an unambiguous notational system and the lack of a systematic way of quantifying linguistic intuitions, or do they represent real disagreements about the acceptability of the structures in question?
The validity problem

- Absolute vs. relative judgments.

Absolute judgments require a decision as to whether (or to what extent) a stimulus has a particular property. People tend to use their own implicit reference point.

Relative judgments require a comparison between two or more stimuli with respect to a particular property.

- How many distinctions?
The reliability problem

- Consistency across and within subjects
- Do intermediate points reflect indeterminancy or real gradience? (do speakers agree in their judgement of intermediate sentences or can they only give reliable judgments for structures at the end points of the acceptability continuum?)
Robustness

No application of robust experimental design factoring out extragrammatical factors such as processing complexity, appropriateness of discourse context, frequency, mode of presentation, order of presentation etc, distraction with fillers etc.
Limited ordinal scales

- Limited in their range of values
- How many distinctions?
  - ✓ Completely acceptable and natural
  - ? Acceptable, but perhaps somewhat unnatural
  - ?? Doubtful, but perhaps acceptable
  - ??* Marginal, but not totally unacceptable
  - * Thoroughly unacceptable
  - ** Horrible
Ordinal scales

- Difficult to interpret: what do adjacent points in an ordinal scale mean?
- They cannot capture the relative strength of grammatical violations.
- Precise differences in acceptability between sentences.
ME in psychophysics

- Can human subjects make reliable proportional judgements of physical stimuli? (e.g. brightness, length, loudness etc.).
- Stevens (1975): magnitude estimation
**Task**: subjects are required to associate a numerical judgement with a physical stimulus; once the initial stimulus, or modulus, is presented and a number associated with it, subjects are asked to assign to each successive stimulus a number reflecting the relationship between each stimulus and the modulus.
No restriction on number values means that subjects can indicate as many distinctions as they perceive.

Since a ratio scale is involved, numerical differences between stimuli reflect differences in impressions.

Scaling is not about absolute accuracy of judgements, but about the relative relationships between judgements of stimuli of different intensities.

In psychophysics ME numerical values can be directly compared with measures of physical stimuli giving rise to impressions.
Stevens’ Power Law: Equal ratios on the physical dimensions give rise to equal ratios of judgements (e.g. in judgements of line length, doubling physical line length doubles subjective line length; in judgements of brightness, every time the stimulus energy doubles, the subjective brightness becomes 1.5 times larger).
Are acceptability judgments like any other kind of human judgment?

Position 1: no, acceptability judgments derive from a special cognitive faculty characterized by properties that are not shared by other kinds of behaviour.

Position 2: yes, acceptability judgments obey the same constraints as any other kind of human judgment.
From psychophysics to linguistics

- **Problem:** Linguistic acceptability has no obvious physical continuum against which to compare judgements.
- **Solution:** Cross modality matching (Lodge 1981).
- **Bard, Robertson and Sorace 1996:** Cross modality validation study.
An ME experiment

- Instructions
- Training phase with line length
- Practice phase with sentences
- Main experimental phase
ME instructions

- Instructions for practice phase explaining the notion of proportionality using line length.

- Instructions explaining that linguistic acceptability can be judged in the same way as line length.
ME instructions

- Your task is to judge how good or bad each sentence is by assigning a number to it.

- You can use any number that seems appropriate to you. For each sentence after the first, assign a number to show how good or bad that sentence is in proportion to the reference sentence.
ME instructions

➢ For example, if the first sentence was:

(1) cat the mat on sat the

and you gave it a 1, and if the next example:

(2) The dog the bone ate.

seemed 20 times better, you’d give it twenty. If it seems half as good as the reference sentence, give it the number 0.5.
ME instructions

- You can use any range of positive numbers including, if necessary, fractions or decimals.
- There are no “correct” answers, so whatever seems right to you is a valid response.
- We are interested in first impressions, so don’t spend too long thinking about your judgement.
Experimental phase

- Modulus: a sentence of intermediate acceptability
- Experimental items in random order and interspersed with filler items
- Setting time limits to intervals between sentences may reduce the likelihood of prescriptive responses.
Data analysis: normalisation

- Dividing each numerical value by the value assigned by a given subject to the modulus creates a common scale. Analyses are then carried out on log-transformed judgements.
- Advantage: parametric tests
Some examples

(i) *Which friend Thomas has painted a picture of? (INV)

(ii) *Which friend have Thomas painted a picture of? (AGR)

(iii) *Which picture has Thomas painted a picture of her? (RES)

(iv) Which friend has Thomas painted a picture of?
The graph shows the mean acceptability (logs) of various constraint violations. The x-axis represents different constraint violations (none, INV, AGR, RES, INV+AGR), and the y-axis represents the mean acceptability values. The values range from -0.5 to 0.1.
Soft Constraints

- ?Which friend has Thomas painted the picture of? (DEF)
- ?Which picture has Thomas torn up a picture of? (EXIST)
- ?How many friends has Thomas painted a picture of? (REF)
Power law for linguistic stimuli
Gradience as an epiphenomenon of interfaces or as a true feature of grammars.

Hard vs. soft constraints
Hard vs. Soft Constraints
(Sorace and Keller 2002)

- Strong vs. mild unacceptability.
  But how can we decide strong and weak in a continuum?
- Soft constraints interact with discourse/semantic factors while hard constraints do not.
- Hard constraints relate to core grammar while soft ones to the interfaces.
- The type of constraint is stable crosslinguistically.
Some interfaces:
syntax-semantics/discourse (interpretation)
syntax-phonology
lexicon-syntax
syntax-processing
Discourse-syntax-phonology

(i) I Maria tha diavasi to vivlio
    the-nom Maria will read-3sg the book
    Maria will read the book
(ii) To vivlio tha diavasi I Maria
(iii) Tha diavasi I Maria to vivlio

SVO, OVS, VSO

Accent on first or rightmost NP
(Keller and Alexopoulou 2001)
Object focus

What did Maria read?

Accent on object NP
Not v-initial
svO, Ovs

Interface Conditions

Information Structure-Phonology:
Accent on the focused constituent.

Information Structure-Syntax: optional movement of the focused constituent to a preverbal position.
All focus: VSO as good as SVO

![Bar chart showing acceptability of word order (no clitic doubling)]
**Object Focus**: accent placement is the strongest cue
Lexicon Syntax Interface
(Keller & Sorace 2003)

Sorace 2000: Auxiliary Selection Hierarchy

**BE**
- change of location
- change of state
- continuation of state
- existence of state
- uncontrolled process
- controlled process-motional

**HAVE**
- controlled process-unmotional
Hierarchy of unaccusative ↔ unergative verbs

The two types have distinct syntax:
Unaccusatives lack an internal subject.
Unergatives have an internal subject.
Change of location
1. Der Gefangene ist/*hat schnell entkommen
   the prisoner is/*has quickly escaped

Change of State
2. Das Kind ist/*hat schnell verstorben
   the child is/has quickly grown

Continuation of State
3. Der Wanderer ?ist/hat kurz verweilt
   the hiker is/has briefly stayed

Existence of State
4. Das Buch hat/*ist mir gefallen
   the book has/is me pleased
**Uncontrolled Process**
5. Die Frau hat/*ist angstvoll gezittert
   the woman has/is fearfully shivered

**Controlled Process (Motional)**
6. Die Frau ist/?hat schnell geschwommen
   the woman is/has rapidly swam

**Controlled Process (Non-Motional)**
7. Die lehrerin hat/*ist dauernd geredet
   the teacher has/is continuously talked
Superiority in English (Featherston 2004)

1a. Who bought what?
   b. ?*What did who buy?
   c. ?What did which boy buy?

2. The dentist showed the toothpaste to the patient.
   a. Who showed the toothpaste to who?
   b. Who did the dentist show what?
   d. Who did the dentist show what to?
D-linking
3 a. Which dentist showed the toothpaste to who?
   b. Which dentist showed the toothpaste to which patient?
   c. Which patient did the dentist show what?
   d. Which patient did the dentist show which toothpaste?
   e. Who did the dentist show what to?
   f. Which patient did the dentist show what to?
   g. Which patient did the dentist show which toothpaste to?
English: Superiority
English: superiority and d-linking
If interfaces are of the same nature crosslinguistically, then variation in this domain can only be quantitative in nature, confined to variation in the magnitude of otherwise identical principles (ultimately reducible to structural differences between grammars).
GERMAN Superiority

![Graph showing 95% CI Judgements for different variables such as wh-subj, wh-DO, wx-DO, wh-IO, wx-IO, and wh-adj. The graph includes error bars indicating the confidence intervals.](image-url)
German superiority and d-linking
Crosslinguistic comparisons

Does the star have the same meaning crosslinguistically?

(i) *Who did you meet him?
(ii) *pion ton idhes
    who-acc him saw-2sg
Intrusive Resumption

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek

Resumption and strong islands in German, English and Greek
Can we reliably distinguish hard from soft constraints?

- Resumption in English and German induces a strong violation (hard constraint) but may improve in certain contexts (soft constraint).
Resumption and weak islands in German, English and Greek:
Accent placement in Greek:

- Violations of accent placement lead to a categorical division of data.
- But, there IS interaction with context.
- Ultimately qualitative criteria appear to be more important than quantitative for the distinction between hard and soft constraints.
- Is the type of constraint constant crosslinguistically?
Gradience does appear to be a property of interface phenomena.

It is not obvious that a distinction between hard and soft constraints is definable in a continuum of judgements.

If gradience is an epiphenomenon of interface phenomena, what about gradience in “core” grammar principles?
WebExp2


http://www.webexp.info/
WebExp: main features

- Different experimental paradigms supported.
- Automatic authentication of subjects’ details and e-mail address.
- Automatic randomisation of experimental materials for each subject.
- Time responses recorded; automatic checks can be carried out on both onset and completion times.
- Data response storage easily processed by standard statistics packages.
- Validation studies indicate high correlation with lab-based data.
Conclusion

- Magnitude Estimation can be used successfully by naïve subjects to judge linguistic acceptability.
- ME results are highly replicable between experiments.
- ME allows us to investigate sophisticated hypotheses derived from linguistic theory and can be a useful tool for reliable comparative work.
But

- We need to better understand the methodology to interpret the results.
- Are judgements really like physical stimuli? (e.g. there is no zero point).
- Is there a power law?
- Is there no difference between categorical and gradient judgements?
Sample References on Magnitude Estimation and gradient acceptability judgements

• Sprouse J 2007 A program for experimental syntax; finding the relationship between acceptability and grammatical knowledge, PhD thesis, U Maryland.
• Sorace A, Acceptability judgements and magnitude estimation in experimental linguistic research, lectures at EMLAR 2006.