THE COGNITIVE COMPLEXITY OF SPLIT INTRANSITIVITY: ACCEPTABILITY JUDGMENTS, ONLINE PROCESSING, AND IMPLICATIONS FOR LANGUAGE

Antonella Sorace
University of Edinburgh
antonella@ling.ed.ac.uk

São Paulo, 21 August 2013

The Auxiliary Selection Hierarchy
(Sorace 2000, 2004, in press)

<table>
<thead>
<tr>
<th>Primary dimension</th>
<th>Secondary dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>+TELIC (KW: BE)</td>
<td>-TELIC (KW: HAVE)</td>
</tr>
<tr>
<td>-TELIC (KW: HAVE)</td>
<td>+TELIC (KW: BE)</td>
</tr>
</tbody>
</table>

Data: what kind?

- Offline vs. online data.
- Acceptability judgments are offline data because they don’t offer any insights on how the human processor constructs a representation of a sentence, which results in a judgment.
- Online data (self-paced reading, eye-tracking, priming etc.) allow the examination of ongoing processing while the sentence unfolds.

A brief journey into magnitude estimation methods

Some representative quotes, to start with…

- “The world of linguistics has changed, changed utterly” (Featherston 2007)
Hayes (1997: 15):
“Linguistics at present is not hard enough; we are not presenting our theories with sufficient demands to distinguish which ones are true. The task of analyzing data with gradient well-formedness puts a theory to a stiffer test.”

Kluender (1998):
“Our job as linguists can be substantially simplified if we try to ferret out those linguistic phenomena that should be handled nonsyntactically. If we can spare our syntactic theory the burden of having to account for such phenomena, it will obviously result in a simpler grammar”.

Weskott & Fanselow (2011)
“The recent years have seen the common practice by linguists of basing their theories on intuitive judgments of grammaticality or acceptability come under attack. This attack was spearheaded by a number of publications, one of which was a highly influential article by Bard, Robertson, and Sorace published in 1996 in Language.”

Conventional measurements of linguistic acceptability

- Judgments of linguistic acceptability usually form category / nominal scales (acceptable, *) or limited ordinal (acceptable, ?, *, **) scales.
- These scales require absolute rating judgments, rather than relative ranking judgments.
- Ordinal scales do not provide information about the relative distance between adjacent points on the scale.

- Nominal scales: numbers are labels with no mathematical properties.
- Ordinal scales. Numbers indicate the relative position of items, but not the magnitude of difference (example: Likert scales, preference rankings).
- Interval scales: numbers indicate the magnitude of the differences between items, but there is no absolute zero point (example: attitude scales and opinion scales).
- Ratio scales: numbers indicate magnitude of difference and there is a fixed zero point. Rations can be calculated (example: age, income, costs, etc.)
Disadvantage of conventional scales

Measurements on these scales have several disadvantages:
• They are limited in their range of values
• They are inconsistent in application
• They are not susceptible to analysis via parametric statistics
• They are unsuited to comparisons between effects of different linguistic constraints, to estimates of systematic variability of judgments, etc.
• They are difficult to interpret (what do the middle points on a rating scale mean?)

What these scales can’t capture

• The relative strength of syntactic violations.
• Gradience within the domain of application of syntactic principles.
• Developmental optionality.
(among many other things…)

In these cases, we want to measure:

• The precise differences in acceptability between sentences.
• The strength of preferences expressed by subjects for one sentence over another

Is there a way of doing this?

• We thought so when we (Bard, Robertson and Sorace 1996) first proposed the use of magnitude estimation in linguistics, based on the results of my 1993 PhD thesis.

ME in psychophysics

• Psychophysics: the scientific study of the relation between stimulus and sensation.
• Magnitude estimation is an experimental technique used to determine quickly and easily how much of a given sensation a person is having.
• S.S. Stevens was the first experimenter to suggest using magnitude estimations to scale sensation in a quantitative way.

• In a magnitude estimation experiment subjects are presented with a standard stimulus (a modulus) and are asked to express the magnitude by a number.
• The subjects are then presented with a series of stimuli that vary in intensity and are asked to assign each of the stimuli a number relative to the standard stimulus.
Subjects assign a number:
• to first stimulus (the *modulus*), to reflect magnitude of pertinent characteristics (length, loudness, brightness, etc).
• to each successive stimulus to indicate apparent magnitude relative to the first.

**Scaling**
• Scaling in ME is **not** about absolute accuracy of judgments;
• Scaling is about the relative relationships between judgments of stimuli of different intensities.

**Different modalities**
• The numerical modality is the most common but other modalities are possible (e.g. line length).
• Other modalities can be more user-friendly particularly if you are testing people who (think they) are numerically-challenged.

**How can you be sure subjects understand how to perform magnitude estimations?**
• Many magnitude estimation experiments use a control condition in which subjects are asked to perform magnitude estimations of the length of a line.
• Magnitude estimations of line length have been shown to be proportional to the actual length of the lines.

**Advantages of ME for physical dimensions**
• ME provides measurements of subjective impressions on a numerical scale which can be plotted against the objective measure of the physical stimuli giving rise to the impressions.
• It does not restrict the number of values which can be used.
• Linear regression of estimates against physical measures in log-log coordinates produces a straight line with a slope characteristics of the physical property being assessed: equal ratios on the physical dimension give rise to equal ratios of judgments (Stevens’ “Power Law”).
The Power Law

- The magnitude of sensation varies as the intensity of the physical stimulus raised to some power \( m \)
- \( S = \text{sensation} \), \( a = \text{constant} \), \( I = \text{intensity} \), \( m = \text{exponent for a particular sensation} \)
- when plotted on log-log axes, the power law plots as a straight line with a slope of the exponent.

\[ S = k I^a \]

Power law

\( a < 1 \): large increases in intensity produce small increases in sensation

\( a > 1 \): small increases in intensity produce large increases in sensation

\( a = 1 \): linear relationship; change in intensity corresponds to sensation change

What about linguistic acceptability?

- Unlike loudness, brightness, length etc., linguistic acceptability has no obvious ‘physical’ continuum to plot against the informants’ impressions.
- There is no zero-point of acceptability.

A “Power Law” for linguistic acceptability?

- Keller (2003) has recently argued that a power law of the same kind as that obtained in psychophysics can be derived by plotting estimated linguistic acceptability against the number of linguistic constraints violated in the stimuli.

Extensions to non-physical domains

- Magnitude estimation has been adapted to judging psycho-social continua with no objective metric: prestige of occupations, support for political policies, etc.
- Magnitude estimation was used on acceptability judgments for the first time by Sorace (1993) not to plot any function, but simply to compare the results with those obtained using more familiar techniques.
Applying ME to linguistic acceptability: advantages

- ME yields interval scales, which allow the application of parametric statistics.
- Mathematical operations can be applied to the estimates, allowing:
  - a direct indication of the speaker’s ability to discriminate between grammatical and ungrammatical sentences;
  - a direct measure of the strength of speakers’ preferences.

Applying ME to linguistic acceptability: advantages

- Informants are enabled to express their intuitions without any restrictions of the judgment scale.
- They are asked to provide purely comparative judgments: these are relative both to a reference item and the individual subject’s own previous judgments;
- At no point is an absolute criterion of grammaticality applied.
- The subjects themselves fix the value of the reference item relative to which subsequent judgments are made.

- The scale used by informants is open-ended and has no minimum division: subjects can always add a further highest score or produce an additional intermediate rating.
- The result is that subjects are able to produce judgments which distinguish all and only the differences they perceive.

Now let’s turn to experimental data.

Gradience in Italian auxiliary selection for unaccusative verbs
(Bard, Robertson & Sorace, 1996)

- Native Italian speakers have most determinate intuitions for core verbs.
- Non-native intuitions gradually approximate the native pattern, starting from core verbs.

Gradience in German auxiliary selection
(Keller & Sorace, 2003)

- Native German speakers have most determinate intuitions for core verbs.
- The telicity effects of prefixes are very clear.
- There is more indeterminacy in the middle of the ASH.
Auxiliary selection across German varieties

• Core verbs elicit the least variation across northern and southern varieties.
• Core verbs elicit the strongest preferences.

Auxiliary selection in Dutch (Sorace & Vonk, 1998)

• Native Dutch speakers have strongest preferences for either zijn or hebben with core verbs.
• Agentivity matters only for verbs at the unergative end of the ASH: agentive subjects elicit stronger preferences for hebben.

Quantifier floating with unergative verbs in Japanese (Sorace & Shomura 2001)

• QF is rejected most strongly with core verbs
• QF is judged as weakly ungrammatical with non-core verbs.

Online processing measures

• Most of the experimental data come from acceptability judgment tests (= the final product of processing).
• Online methods do not rely on explicit responses to stimuli on which the speaker has conscious control.
• These methods provide a more direct picture of implicit knowledge that cannot be consciously manipulated.

As we say in Bard, Frenck-Mestre and Sorace (2010)...

• AJ data don’t provide information on the source of the ASH effects.
• In particular, they tell us little about the processing events that lead to the expression of judgments about the acceptability of one auxiliary or the other with any given verb type and in any given context.

Is it just a question of lexical associations?

• Is the gradience given by the strength of relationships between entries in the mental lexicon?
• Is the ASH gradience the effect of lexical associations?
Lexemes

Priming condition:
- Screen 1: è (is)
- Screen 2: caduto (fallen-PP)
Required response: caduto (fallen-PP)

Cueing condition:
- Screen 1: caduto (fallen-PP)
- Screen 2: è (is)
Required response: caduto (fallen-PP)

Lemmas

Priming condition:
- Screen 1: è
- Screen 2: cadere (fall-INF)
Required response: caduto (fallen-PP)

Cueing condition:
- Screen 1: cadere (fall-INF)
- Screen 2: è
Required response: caduto (fallen-PP)

Experiment 2: Primed/cued Word Naming for Participles (Bard et al 2010)

Experiment 3: Primed/cued Word Production for Participles

- ASI effect for RT only in primed condition only for unaccusatives
  \( F(1,11) = 6.02, p = .027; NK at p = .015 \)
  \( F(1,14) = 6.23, p = .026 \; NK at p = .015 \)
**Experiment 3:**
**Primed/cued Word Production for Particles**

The only effect is a relationship between the auxiliary and the core examples of the verb class using that auxiliary.

**Unaccusative**
- Core: \( \dot{e}/ha \) cadere-caduto
- Periphery: \( \dot{e}/ha \) durare-durato

**Unergative**
- Core: \( *\dot{e}/ha \) riflettere-riflettuto
- Periphery: \( *\dot{e}/ha \) cedere-ceduto

- E primes core (but not peripheral) unaccusatives.
- A doesn’t prime unergatives (either core or peripheral).
- Question: why is there a privileged relationship only between E and core unaccusatives?

**Eye-tracking methods**

- Are rigidity and flexibility at the syntax–lexicon interface reflected by ocular evidence?
- Do verbs with the wrong auxiliary require longer reading times compared to verbs with the right auxiliary?
- Is the difference between right and wrong auxiliary greater for core verbs compared to non-core verbs?

**Eye-tracking experiments on Italian** (Bard, Frenck-Mestre & Sorace 2010; Sorace & Vernice, forthcoming)

- \( *\text{Alla festa*il miliardario}*ha/è entrato*da solo*nelle sala*\n- \( *\text{Alla festa*il miliardario}*ha/è rimasto*da solo*nelle sala*\n
At the party the millionaire entered/remained alone in the room

- \( *\text{A quella vista*il codardo}*ha/è urlato*per lo*spavento*\n- \( *\text{A quella vista*il codardo}*ha/è trasalito*per lo*spavento*\n
At that sight the coward shouted/jumped in fright

**Eye movement measurements**

1 2 3 4 5
(*Alla festa*) il miliardario*ha/è entrato*da solo
*nella sala*

- Regions of interest: 2 (subject), 3 (V+Aux), 4 (first portion of PP immediately following the participle), and 5 (remainder of PP)
- First-pass reading (the total reading time spent on the segment before exiting that segment)
- Total reading time (sum of all fixations on a segment, including regressive eye movements).

**What the eye movement data show**

- Total reading times with core verbs (both unaccusative and unergative) are significantly shorter for correct auxiliaries than for incorrect auxiliaries.
- No difference in reading times for correct and incorrect auxiliary with non-core verbs.
Unergatives/total reading times

Unaccusatives/total reading times

Signatures of gradience

- Spill-over effects for non-core verbs, especially for non-core unergatives.
- Interpretable as non-commitment of the processor on auxiliaries with non-core verbs until the rest of the sentence is encountered.

Different salience of subjects

- Auxiliary violations with unergatives trigger **more fixations on the subject** than auxiliary violations with unaccusatives.
- The subject of unergatives is more salient for auxiliary selection than the subject of unaccusatives.
Hypotheses

• Core unergative verbs should lead to a more salient representation of the subject argument because of its agentive nature and its syntactic stability \( \rightarrow \) better implicit memory performance.
• Graded levels of subject salience depending on place of the verb on ASH.

Hypotheses

• The subject of unergative verbs should be more memorable than the subject of unaccusative verbs.
• The subject of core unergative verbs should be more memorable than the subject of non-core unergative verbs.

Eye movement data: a short summary

• Auxiliary selection violations with core verbs cause more processing disruption than violations with non-core verbs.
• Auxiliary violations with non-core verbs cause more extended processing disruption than those to core verbs because non-core verbs depend on compositional factors beyond the AUX + V combination.

EEG measures of implicit memory for the subject of intransitive verbs

• ERPs: scalp-recorded changes in electrophysiological activity in the cortex which are time-locked to external events.
• Reflexes of split intransitivity on implicit argument salience measured in terms of electrophysiological activity.

Sorace, Filiaci & Demiral 2012

• Core and non-core unaccusative and unergative verbs in the present tense (no auxiliaries) were included.
• Only grammatical sentences:
  Da secoli un vampiro appare verso la mezzanotte
  For centuries a vampire has appeared around midnight
  Nel film il pirata viaggia per i sette mari
  In the film the pirate travels the seven seas
Method / core unaccusative

Da secoli un vampiro appare verso la mezzanotte
For centuries a vampire has appeared around midnight

Does this match the referent of the sentence you have just read? MATCH / MISMATCH

Method / core unergative

Nel film il pirata viaggia per i sette mari
In the film the pirate travels the seven seas

Does this match the referent of the sentence you have just read? MATCH / MISMATCH

Effects in the EEG signal

- Brain responses in the theta frequency band (4-7Hz) of the human EEG are involved in memory performance with respect to load on working memory resources (higher theta -> more retrieval effort).
- The theta activity for the argument of core unaccusative verbs is higher than the evoked theta activity for the argument of unergative verbs (p<.05).
- No difference in the theta activity for non-core verbs.

What’s the status of non-core verbs?

- Are they flexible because they are syntactically or semantically indeterminate?
- Or are they flexible because they have a basic syntactic behavior but are also compatible with more than one syntactic projection?

ERP experiments on German

(Roehm & Sorace 2008; Roehm, Sorace & Bornkessel-Schlesewsky, 2010, in press)

- We measured the ERP signal in response to auxiliary violations in German with verbs encoding telicity in different ways.

ERP questions

- Is the distinction between rigid and flexible syntax-lexicon mappings reflected in patterns of neurophysiological activity?
- How do the electrophysiological events associated with the processing of syntax-lexicon mappings correlate with sentence acceptability?
Exp. 1: telicity

Aim was to measure ERP differences between:
• (a) lexically telic verbs ("core unaccusative", e.g. abreisen 'leave') and atelic agentive verbs ("core unergative", e.g. arbeiten 'work'),
• (b) telicity inherently encoded in the verb's argument structure (as in ankomen 'arrive') and telicity morphologically induced by the presence of a prefix (as in verrosten 'rust' compared to rosten).

Exp. 2: statives and animacy

Aim was to measure ERP differences between:
• animate vs. inanimate subjects (+/- volitional initiators) with both:
  (a) verbs denoting a telic change (e.g. abreisen 'leave') those denoting a state (e.g. liegen 'be situated')
  (b) verbs denoting controlled (run) vs. uncontrolled process (buzz).

Participants

• Exp.1: 33 right-handed native speakers of German (17 female) between 20-30 years of age (mean age: 23.97)
• Exp.2: 30 right-handed native speakers of German (16 female), age range: 20-32 (mean age: 24.8)

Materials

• Exp.1: 4 verb classes, 8 different verbs, 10 different sentence contexts per verb; all sentences of the form NP/AUXILIARY/ADVERB/PAST PARTICIPLE, half with the AUX sein, half with the AUX haben.
• Exp.2: 4 verb classes, 24 different verbs, 4 critical sentence conditions (+/-anim, haben/sein). All sentences were presented as a subordinate clause which followed a short matrix clause and had the form DASS/NP/ADVERB/PAST PARTICIPLE / AUXILIARY.
What the ERP data show / Exp.1

- Auxiliary selection violations with core *haben* and *sein* verbs both elicited a biphasic N400/P600 complex pattern.
- The pattern is more pronounced for core verbs than for non-core verbs.
- This pattern has been reported in the ERP literature as associated with syntactic violations that induce a simultaneous semantic processing difficulty (Rösler, Putz, Friederici and Hahne 1993; Ainsworth-Darnell, Shulman and Boland 1998; Hagoort 2003).

N400-P600

- Prefixed change of state verbs also show a N400-P600 biphasic component: however, the N400 is smaller for prefixed change of state verbs than for change of location verbs, suggesting a difference between inherent and compositional telicity.
- Unprefixed change of state verbs showed no effect, consistent with the interpretation of these verbs as indeterminate with respect to telicity.

What the ERP data show / Exp.2

- Stative verbs and verbs of uncontrolled process elicited a N400/P600 for *sein*.
- Animacy effects (late negativity) were obtained only for stative verbs and verbs of uncontrolled process.
- Verbs of controlled motion elicited only a N400 for *haben* indicating that both auxiliaries are acceptable with this verb class but with associated interpretive differences.
Telicity

Aim was to measure ERP differences between:
• (a) lexically specified telic verbs ("core unaccusative", e.g. abreisen ‘leave’) and atelic agentive verbs ("core unergative", e.g. arbeiten ‘work’),
• (b) telicity inherently encoded in the verb’s argument structure (as in ankennen ‘arrive’) and telicity morphologically induced by the presence of a prefix (as in verrosten ‘rust’ compared to rosten).

Materials

• All sentences of the form NP/AUXILIARY/ADVERB/PAST PARTICIPLE

Example:
Die Bergsteiger ist/that vorsichtig aufgestiegen
The mountaineer is/has carefully climbed

Sample sentences

<table>
<thead>
<tr>
<th>Condition</th>
<th>Verb class</th>
<th>Example</th>
<th>HABEN</th>
<th>SEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-LOC</td>
<td>Die Bergsteiger ist/that vorsichtig aufgestiegen</td>
<td>1.9 (0.3)</td>
<td>93.5 (3.8)</td>
<td></td>
</tr>
<tr>
<td>CH-STATE-(CN)</td>
<td>Die Dose ist/that vorsichtig aufgestiegen</td>
<td>61.0 (7.0)</td>
<td>65.0 (10.0)</td>
<td></td>
</tr>
<tr>
<td>CH-STATE-(PRE)</td>
<td>Der Auto ist/that vorsichtig aufgestiegen</td>
<td>4.0 (4.0)</td>
<td>88.6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>CON-PROC</td>
<td>Der Lehrer ist/that vorsichtig aufgestiegen</td>
<td>94.2 (5.7)</td>
<td>1.2 (2.0)</td>
<td></td>
</tr>
</tbody>
</table>

Example sentences

What the ERP data show

• Auxiliary selection violations with core haben and sein verbs both elicited a biphasic N400/P600 complex waveform.

• Unprefixed change of state verbs elicit no such effect – there is no difference between haben and sein.

N400 signature of compositional telicity

• Prefixed change of state verbs also show a N400-P600 biphasic component, but the amplitude of the N400 is smaller and the onset is delayed for auxiliary violations of prefixed verbs than for auxiliary violations change of location verbs, indicating a difference between inherent and compositional telicity.

• Additional computational effort required by prefixed verbs.
Correlations with acceptability

- For unprefixed flexible verbs, average ERPs did not show differential effects of auxiliary choice.
- However, additional mixed-effects analyses for these verbs revealed interesting correlations between the amplitudes of N400 and P600 waves and individual acceptability ratings for HAVE-sentences.

Negative correlation between N400 amplitude and acceptability

- Some individuals rate HAVE with unprefixed change of state verbs as acceptable, other individuals rate it as unacceptable.
- Smaller N400 for HAVE correlates with higher acceptability rates for these sentences.
- More pronounced N400 correlates with lower acceptability rates.

Two qualitatively distinct late positivities

- The P600 reflects the categorisation of the sentence as ill-formed, hence it correlates negatively with the acceptability rating for that sentence (as seen before).
- The P600 also reflects the process of coercing a telic verb into an activity reading in order to render it compatible with the auxiliary HAVE, hence it correlates positively with the acceptability rating for that sentence.

Positive correlation between P600 amplitude and acceptability

- Larger ERP amplitudes in the HABEN condition for high acceptability ratings.
- ERPs tend to be more positive for those individual subjects who show a higher acceptability for the unprefixed change of state verbs with HABEN.

Signature of enrichment and aspectual coercion

- An enrichment process of aspectual coercion is necessary in order for the change of state to be interpreted as a process rather than as a telic change.
- If this process of enrichment is successful, the acceptability of CH-STATE-(UN) verbs with HAVE is higher and there is a more pronounced positivity.

Individual differences in aspectual coercion

- Some participants may have a propensity for aspectual enrichment processes, which leads them to consider these verbs more acceptable with HAVE in comparison to participants with a lower tendency to enrich.
Revisiting flexible unprefixed CH-STATE verbs

• Is it the case that CH-STATE-UN verbs are truly indeterminate for HABEN or SEIN (= that both auxiliaries fulfill the processing system’s expectations for this particular verb class)?
  • No.

References / 1


Conclusions

• We have consistent evidence of the difference between verbs at the extremes of the ASH/SIH.
• So we have evidence both for the syntactic split between intransitive verbs and for the gradient effects induced by the semantics of the verb.
• Tomorrow we’ll try to make some sense of these patterns.

References / 2

Roehm, D., Sorace, A. and Bornkessel-Schlesewsky, I. submitted. ERP signatures of auxiliary selection in German.


References / 3


