

Semantic and pragmatic annotation

Volker Gast

Friedrich-Schiller-Universität Jena

31 July, 2019

Semantic and pragmatic annotation

- Conceptual issues
 - standards of empirical research
 - practical questions
- Case studies
 - adverbials of immediate posteriority (*immediately*)
 - scalar additive operators (*even, so much as*)

Matters of empirical analysis

- The analysis of structural levels requires analytic decisions concerning observable linguistic matter.
- Meaning and function cannot be directly observed.
- Standards of empirical analysis
 - validity (of measurements or classifications)
 - reliability (of measurements or classifications)

Validity and operationalization

- If a category cannot be observed, it has to be operationalized.
- Operationalization is a procedure that renders an unobservable property/category observable.
- It is a matter of validity; do we really measure/categorize what we intend/claim to measure/categorize?
- An example from economy:
 - innovativity of a region, measured as the number of patent applications submitted
 - degrees of cooperativity, measured as the number of **joint** patent applications submitted

Reliability

- Reliability is the degree to which a measurement/classification is invariant across repeated measurements/classifications.
- Important procedure in linguistic annotation: determine inter-annotator reliability.
- Coding scheme: a system of categories into which observations are classified.
- Coding guidelines: conventions followed by the annotators.
- Ways of measuring inter-annotator reliability
 - annotation via a web interface
 - annotation in GraphAnno

Adverbials of immediate posteriority

- Objectives: Determine differences in meaning/function/distribution, across languages (English, German, Spanish, Italian, French) and registers (Europarl, OpenSubtitles).
 - Reflections of such differences in translation.
- (1) Je vais l'envoyer tout de suite [immédiatement].
“I'll send it in a minute / immediately.” [OS]
 - (2) Bin gleich [sofort] wieder da.
“I'll be back right now / immediately.” [OS]

Variables of annotation

- Properties of adverbials of immediate posteriority:
 - the reference point
 - intentionality
 - the “controller” of the action
 - modality and illocutionary force
 - information structure

The annotation interface

id	original	translation							
1	<p>Juste après 9 h. 30, le socialiste espagnol, Manuel Medina Ortega, quitte l'hémicycle.</p> <p>Je vous en laisse juge. «Juste après 9 h. 30, le socialiste espagnol, Manuel Medina Ortega, quitte l'hémicycle. Un instant plus tard, il arrive au bureau des chauffeurs.</p>	<p>Gleich nach 9.30 Uhr verläßt der spanische Sozialist Manuel Medina Ortega den Plenarsaal.</p> <p>Ich lasse Sie selbst entscheiden. Gleich nach 9.30 Uhr verläßt der spanische Sozialist Manuel Medina Ortega den Plenarsaal. Kurz danach kommt er im Fahrerbüro an.</p>							
<p>Can the example be annotated?</p> <p><input checked="" type="radio"/> y</p> <p><input type="radio"/> n</p>									
orig	<p>aller</p> <p><input type="radio"/> y</p> <p><input type="radio"/> n</p>	<p>mod-fr</p> <p>indicatif</p> <p>subjonctif</p> <p>conditionel</p> <p>imperatif</p> <p>infinitif</p> <p>participle-pres</p> <p>participle-passe</p>	<p>temp-fr</p> <p>pres</p> <p>imparfait</p> <p>passe-compose</p> <p>passe</p> <p>plus-que-parfait</p> <p>futur-simple</p> <p>futur-compose</p>	<p>modal</p> <p>nicht-modal/assertiv</p> <p>Aufforderung/Wunsch</p> <p>Versprechen/Drohung</p> <p>Warnung</p> <p>Möglichkeit/Fähigkeit</p> <p>keine_Modaltitätsmarkierung</p> <p>Vermutung</p>	<p>pers</p> <p>1pl</p> <p>2pl</p> <p>3pl</p> <p>3-belebt-unpersönlich</p> <p>3-belebt-unpezifisch</p> <p>3-unbelebt</p> <p>na</p>	<p>ref</p> <p><input type="radio"/> deic-dir</p> <p><input type="radio"/> deic-phant</p> <p><input type="radio"/> ce</p> <p><input type="radio"/> cz</p> <p><input type="radio"/> na</p>	<p>int</p> <p><input type="radio"/> t</p> <p><input type="radio"/> f</p> <p><input type="radio"/> na</p>	<p>modifiziert</p> <p><input checked="" type="radio"/> t</p> <p><input type="radio"/> f</p> <p><input type="radio"/> na</p>	<p>ip-fr</p> <p>tout-de-suite</p> <p>immédiatement</p> <p>zero</p> <p>aller+inf</p> <p>dès</p> <p>d-emblée</p> <p>tout-à-t-heure</p>
tri	<p>mod-de</p> <p>indikativ</p> <p>konj1</p> <p>konj2</p> <p>infinitiv</p> <p>imperativ</p> <p>partizip1</p> <p>partizip2</p>	<p>temp-de</p> <p>praes</p> <p>praet</p> <p>perf</p> <p>plaperf</p> <p>fut</p> <p>fut2</p> <p>na</p>	<p>modal</p> <p>nicht-modal/assertiv</p> <p>Aufforderung/Wunsch</p> <p>Versprechen/Drohung</p> <p>Warnung</p> <p>Möglichkeit/Fähigkeit</p> <p>keine_Modaltitätsmarkierung</p> <p>Vermutung</p>	<p>pers</p> <p>1sg</p> <p>2sg</p> <p>3sg</p> <p>1pl</p> <p>2pl</p> <p>3pl</p> <p>3-belebt-unpersönlich</p>	<p>ref</p> <p><input type="radio"/> deic-dir</p> <p><input type="radio"/> deic-phant</p> <p><input type="radio"/> ce</p> <p><input type="radio"/> cz</p> <p><input type="radio"/> na</p>	<p>int</p> <p><input type="radio"/> t</p> <p><input type="radio"/> f</p> <p><input type="radio"/> na</p>	<p>modifiziert</p> <p><input checked="" type="radio"/> t</p> <p><input type="radio"/> f</p> <p><input type="radio"/> na</p>		
<input type="button" value="submit data set"/>									

Figure 2: The annotation interface

The reference point

- Adverbials of immediate posteriority can be either chronological or deictic.

(4) deictic reference point

– Kommst du?

– Ja, gleich/sofort/bald/*anschließend/*daraufhin.

“Are you coming? – Yes, in a minute /immediately /soon / *subsequently.”

(5) anaphoric reference point/event

Ich klopfte und er öffnete mir sofort die Tür.

“I knocked and he opened the door immediately.”

Intentionality

- Some adverbials of immediate posteriority are sensitive to the parameter of intentionality.
- (8) Oh, Gott! Es explodiert gleich [/*sofort]! Lauf!
“Oh my God! It’s gonna blow! Run!” [OS]
- (9) – Turner, wieso läuft das Ding noch nicht?
– Bin gerade eingetroffen, Sir. Der Generator ist gleich [/sofort] betriebsbereit.
“– Turner, why is this thing not running? – I only just arrived, Sir. The generator will be operational in a moment.” [OS]

Modality and illocutionary force

- Categories used for annotation:
 - directive/request
 - commissive/threat
 - warning
 - representative:
 - possibility/ability
 - inference/supposition
 - necessity
 - non-modal

Results: OpenSubtitles

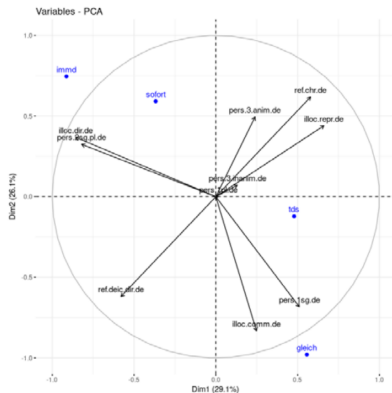


Figure 15: Principal Component Analysis: All predictors [OS]

Results: Europarl

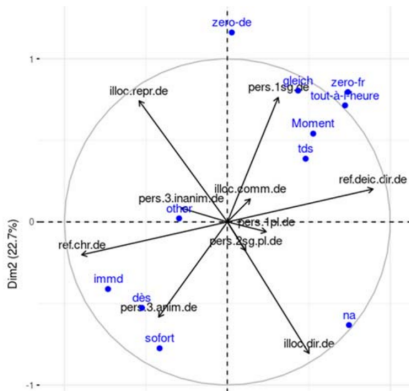


Figure 16: Principal Component Analysis: All predictors [EP]

Results

- The semantic/distributional space is partitioned very differently in French and German.
- Differences in markedness/specificity:
 - French *tout de suite* is more or less generic and covers a broad range of contexts, whereas *immédiatement* is rather specialized for directive contexts.
 - German *sofort* is more generic than *gleich*, which is associated with commissive speech acts and the first person.

Even as a scalar operator

- (1) Bill Nighy is very famous.
(a) *Even* the Queen congratulated him on his birthday.

(b) *Even* $\left\{ \begin{array}{l} \text{the Queen} \\ \uparrow \text{his neighbour} \\ \uparrow\uparrow \text{his mother} \end{array} \right\}$ congratulated him on his birthday.

- (2) Bill is very happy.
 $\uparrow\uparrow$ *Even* his mother congratulated him on his birthday.

Even under negation

(3) Bill is very unhappy.

(a) Not *even* his mother congratulated him.

(b) Not *even* $\left\{ \begin{array}{l} \text{his mother} \\ \text{†his neighbour} \\ \text{††the Queen} \end{array} \right\}$ congratulated him.

Even and so much as

- (4) If you *so much as* [*~ even*] parked on a yellow line they stuffed a mortgage application under your windscreen wipers. [BNC, SMA 4]
- (5) Can the calculation of 165 deaths per one million rems be applied to all age groups—or *even* [*?so much as*] any? [BNC, EVEN 1065]
- (6) That awful thing that so many groups get themselves involved in, when they're on a plane and they do a gig and they don't *even* [*?so much as*] know what city they're in, he'd manage to avoid. [BNC, EVEN 285]

Parameters of variation

- Properties of scalar additive operators:
 - The syntactic category of the co-constituent of the operator;
 - the syntactic category of the focus;
 - the type of downward entailing operator;
 - the relation ordering the set of focus alternatives;
 - the presence or absence of focus alternatives in the clausal environment.

The category of the co-constituent and the focus

- The co-constituent and the focus are not necessarily identical.
 - (8) Relatively few voters read *even* [_{NP} the [_A best-selling]_F papers]. [BNC, EVEN 139]
 - (9) It is not a problem that is [_V solved]_F, or *even* [_V touched]_F, by another 10s. [BNC, EVEN 166]

The category of the co-constituent

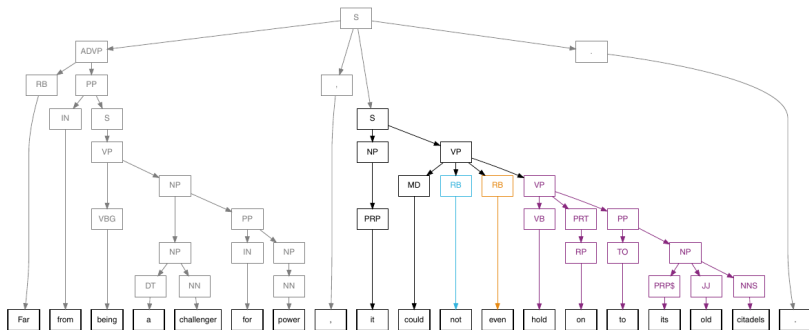


Figure 2. Example (2) with sentence constituents classified

The downward entailing operator

(10) negation with *not*

Danilov, The Voice observed, had probably **not even** begun to contemplate his murder when Dostoevsky was shaping Raskolnikov's. [BNC, EVEN 71]

(11) *without*-PP

There was another rustle of branches as the buffalo ran off **without so much as** another snort. [BNC, SMA 7]

(12) conditional

The new Lady Woodleigh looked as if she might take her riding-crop to him **if he so much as** uttered another word. [BNC, SMA 5]

The downward entailing operator

(13) *before*-clause

It reminded me of all I disliked so much in the United States, of being called Ray **before** *even* shaking hands. [BNC, EVEN 94]

(14) superordinate lexical trigger (here, *unwise*)

Biggs is of the opinion that Mason would be unlikely to survive more than a couple of rounds against the world heavyweight champion and at this stage it would be **unwise** to *even* think of him as a genuine contender. [BNC, EVEN 10]

(15) negative determiner

No one *so much as* raised an eyebrow in their direction. [BNC, SMA 22]

The ordering of alternatives

(16) foci are lexically ordered

“If you’ll pardon the correction, not *so much as* [a million]_F,” said one of the lady lodgers. [BNC, SMA 38]

(17) not lexically ordered

She was not *even* [inarticulate]_F in the sense that she could express her own feelings convincingly. [BNC, EVEN 87]

The presence of overt focus alternatives

(18) focus alternative present

Most stable-lads would have counted themselves lucky *even* to [get a ride]_F let alone to [win a race]_F. [BNC, EVEN 211]

(19) focus alternative absent/implicit

[Picked him up at Imperial College. I gave them a three-hour lecture on the basic principles of stochastics, he said.]

Some composers today don't *even* understand [the simple calculus]_F, he said. [BNC, EVEN 16]

A fully annotated sentence

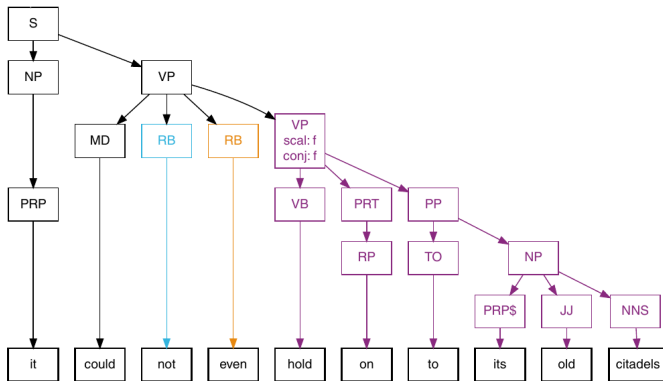
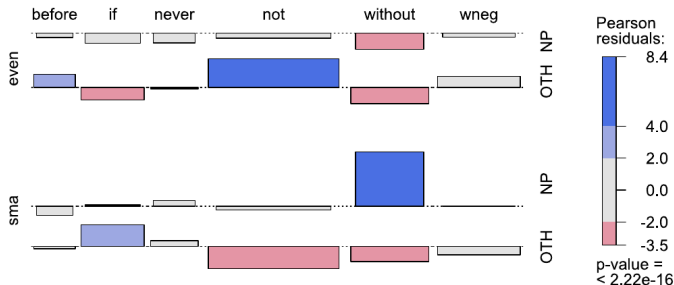


Figure 3. Fully annotated sentence

Quantitative results

- Two variables turned out to be significant predictors distinguishing *even* and *so much as*.



Results in prose

- *So much as* is particularly common in combination with nominal co-constituents, in the scope of *without*, and it is significantly overrepresented in combination with other co-constituents in conditionals.
- *Even* is typically found in the scope of *not* and *before*, with non-nominal foci.

Typical examples

(34) *even* under direct clause negation

But nowadays they call on new and brasher saints, whom St Margaret had not *even* met – Saints Epidura and Psychophilaxia. [BNC, EVEN 20]

(35) *so much as* in conditional

If you *so much as* parked on a yellow line they stuffed a mortgage application under your windscreen wipers. (= [4])

(36) *so much as* in *without*-PP

There was another rustle of branches as the buffalo ran off without *so much as* another snort. (= [11])

An explanation?

- Inertia in language: Properties of earlier language stages are preserved.
- *Even* is younger than *so much as* and came to be used in downward entailing contexts under negation first, sometimes reinforcing *so much as*.
- *So much as* is semantically closer to downward entailing contexts without a negation, as a measure function.

(37) All which abuses, if those acute philosophers did not promote, yet they were never able to overcome; nay, *even not so much as* King Oberon and his invisible army. (Thomas Sprat, *The History of the Royal Society*, quoted from Gast and van der Auwera [2011, 43])

(39) Expansion and duration have this further agreement, that, though they are both considered by us as having parts, yet their parts are not separable one from another, *no not even* in thought. (J. Locke, *An Essay concerning Human Understanding*, XV, 10; 1690; quoted from Gast and van der Auwera [2011, 43])

Summary

- Semantic annotation and empirical standards
- Important concepts:
 - validity
 - reliability