Doing corpus-based typology:
Concessivity in cross-linguistic perspective

Volker Gast

26 Jan 2016

Corpus linguistics and linguistic typology

▶ Complementary roles of corpus linguistics and language typology: Converging evidence from typology and language-specific (or contrastive) corpus studies; e.g. Hawkins' (1994, 2004) 'Performance-Grammar Correspondence Hypothesis':

Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

Corpus-based typology

- Linguistic typology has, for the most part, been based on properties of linguistic systems, rather than 'linguistic output': morphological properties of words (morphological typology), syntactic patterns (word order typology), phoneme inventories (phonological typology), etc.

- Corpus-based typology aims at generalizing over languages not by assigning them to one type – e.g. 'agglutinating, SOV' for Turkish – but by comparing linguistic output ('texts') directly.

Requirements of corpus-based typology

- ▶ Resources, tools, categorizations
- ▶ Technical aspects:
  - ▶ Multilingual corpora (comparable corpora, translation corpora)
  - ▶ An infrastructure for annotating (multilingual as well as monolingual) corpora at various levels, manually – interactively, automatically.
- ▶ Conceptual aspects: Annotation schemes, ideally universally applicable and standardized (making community efforts possible)

This talk

The domain of inquiry: Concessives (and the like)

Concessives in typological perspective

A pilot corpus study: English, German and Spanish

Concessives and their like as markers of 'background adjustment'

▶ Broadly speaking, concessive and similar relations (adversative relations, relations of contrast or antithesis, etc.) have in common that the 'conversational background' is 'adjusted' in some way.

(1)   a. John will have to sit on that robust chair. He's pretty fat.
       b. Well – although he's fat, he is not heavy.

     BG: John is fat → John is heavy
     Adj: John is heavy ⇒ John is not heavy.

(2)   a. John is fat and tall, isn't he?
       b. He's fat but not tall.

     BG: John is fat and John is tall.
     Adj: John is tall ⇒ John is not tall.

Concessives and the like in RST

- Rhetorical Structure Theory makes a basic distinction between *nuclei* and *satellites*.
- Concession and antithesis are satellites in mononuclear relations, contrast holds between two nuclei.
- The difference between concession and antithesis is that concession implies 'violated expectation'.

(3) While [$_S$ John is fat], [$_N$ he is not heavy]. (concession)

(4) While [$_S$ John is fat], [$_N$ he is not tall]. (antithesis)

(5) [$_N$ John is tall], while [$_N$ Mary is slim]. (contrast)

Concessives et al. in the PDTB

- Comparison
  - Contrast
  - Similarity
  - Concession (denial of an expected causal relation)
    - Arg1-as-denier
    - Arg2-as-denier
  - Concession+SpeechAct
    - Arg2-as-denier+SpeechAct

(6) Although he's fat, John is not heavy. (Arg1-as-denier)

(7) John is fat, but he's not heavy. (Arg2-as-denier)

(8) John is slim, or so he claims. (Arg2-as-denier+SpeechAct)

Ted Sanders' classification of DRDs

- Ted Sanders classifies coherence relations in terms of four categories:
  - Polarity (positive, negative)
  - Basic operation (additive, temporal, causal)
  - Source of coherence (objective, subjective)
  - Order (forward, backward)
- Concession is defined as 'negative causal' (either 'forward' or 'backward').
- A distinction can be made between 'subjective' and 'objective' concession.

(9)   a. The glass broke because I dropped it.

        b. The glass didn't break although I dropped it.

(10)   a. John's probably at home, because his car is in the driveway.

        b. John's probably at home, though his car is not in the driveway.

Volker Gast   Concessives and corpus-based typology

Concessive conditionals

- Important category: Concessive conditionals/hypothetical concessives (first argument is not presupposed).
- Concessive conditionals often develop into concessives (cf. König 1991a)

(11)   Although John is drunk he is not impolite.

(12)   Even if John is drunk he is not impolite.

Volker Gast   Concessives and corpus-based typology

Further issues of semantics and pragmatics

- ▶ Semantics:
  - ▶ What exactly is negated/adjusted? A condition or a cause/reason (hard to define)?
  - ▶ Are concessives really the 'dual partners' of causal relations (such as universal and existential quantifiers; cf. König 1991a, Pasch 1992, among others)?
- ▶ Pragmatics:
  - ▶ What status does the negated proposition have – is it a type of presupposition (cf. König 1991b)?
  - ▶ Why can concessives not normally be in focus (or function as the argument of a focus particle)?

(13)   a. Why did you go out?
        b. Because the weather was nice.

(14)   a. In spite of what circumstances did you go out?
        b. #Although it was raining.

Volker Gast Concessives and corpus-based typology

World knowledge and common ground management

- The notion of 'conversational background' subsumes various dimensions, at least:
    - **encyclopaedic knowledge**
    - **presuppositional information** (linguistically marked as such)
    - **topicality** (what is currently under discussion?)

	encyc.	pres.	topic	assertion
adv.	–	–	John's weight and height.	John is fat but not tall
conc.	John is fat $\rightarrow$ John is heavy	John is fat	John's weight	Although John is fat, he is not heavy.
cc	John is drunk $\rightarrow$ John is impolite	–	John's manners	John is not impolite, even if he's drunk.

Volker Gast   Concessives and corpus-based typology

More generally speaking . . .

▶ Background adjustment implies a change in polarity between some proposition in the background and a part of the assertion.

	encyc.	pres.	under disc.	assertion
adversatives	–	–	$a \wedge b$	$a \wedge \neg b$
concessives	$a \rightarrow b$	a	b	$a \wedge \neg b$
conc. conds.	$a \rightarrow b$	–	b	$a \wedge \neg b$

Paradigmatic relations between arguments

- ▶ In specific cases of concessive and adversative relations, a paradigmatic relation holds between the arguments, insofar as they provide information about the same topic (answer the same question under discussion).

- ▶ Such paradigmatic relations may be located at the same level of categorization ('horizontal'), or the arguments may stand in a hierarchical relation to each other ('vertical').

(15) John is fat but not tall.
Topic: John's weight and height, 'horizontal' alternatives

(16) John is nice, though not always.
Topic: John's character, 'vertical' alternatives

(17) I'll go for a walk, although it's raining.
Topic: Speaker's plans, no paradigmatic relation between ARG-1 and ARG-2

Volker Gast   Concessives and corpus-based typology

From implication to likelihood

- What semantic relation is cancelled in concessives: condition, cause/reason?
- Proposal: The background assumption has the status of a 'probabilistic implication'.
- Two possibilities:
  - ARG-1 makes ARG-2 more likely: **pro(ARG-1,ARG-2)**
  - ARG-1 makes ARG-2 less likely: **con(ARG-1,ARG-2)**
- A probabilistic implication may hold between semantic/pragmatic entities at all levels (propositions, speech acts).

Some examples

(18) Relationship between events
The glass didn't break although it fell to the floor.
(The glass's falling increases the likelihood of its breaking.)

(19) John's probably not at home, although his car is not in the driveway.
(John's car in the driveway increases the likelihood of me concluding that he's at home.)

(20) Bring me a beer – though I have to drive ...
(My having to drive increases the likelihood of me not asking you for a beer.)

Reformulating the table . . .

	prob. impl.	pres.	under discussion	assertion
advers.	–	–	a ∧ b	a ∧ ¬b
concessives	pro(a,b)	a	b	a ∧ ¬b
conc. conds.	pro(a,b)	–	b	a ∧ ¬b

Volker Gast   Concessives and corpus-based typology

Reformulating the table . . .

	prob. impl.	pres.	under discussion	assertion
advers.	–	–	a ∧ b	a ∧ ¬b
concessives	pro(a,b)	a	b	a ∧ ¬b
conc. conds.	pro(a,b)	–	b	a ∧ ¬b

▶ Alternatively . . .

	prob. impl.	pres.	under discussion	assertion
advers.	–	–	a ∧ b	a ∧ ¬b
concessives	con(a,¬b)	a	b	a ∧ ¬b
conc. conds.	con(a,¬b)	–	b	a ∧ ¬b

A note on indirect adversatives . . .

- ▶ Adversative connectors such as *but* often make reference to a 'third', implied, element (cf. also Ducrot's analysis of *but*).

(21) a. How's John?
      b. He's nice but lazy.

(22) a. How's John? Is he a potential boyfriend?
      b. He's nice ($\Rightarrow$ yes) but lazy ($\Rightarrow$ no).

	prob. impl.	pres.	topic/QUD	assertion
advers.	–	–	$a \wedge b$	$a \wedge \neg b$
ind. advers.	pro(a,c)	–	$a \wedge b$	$a \wedge \neg b$
	pro(b,c)			

Volker Gast   Concessives and corpus-based typology

Typologies of concessives

- Parameters of variation:
  - Semantics/distribution, formal properties (e.g. Kortmann 1992, Xrakovskij 2012)
  - Internal make-up, historical sources

- **quantifying** expressions ('under all conditions of the type of ARG-1, ARG-2 holds', e.g. Engl. *al-though*)
- **scalar** expression ('even under the most unlikely condition stated in ARG-1, ARG-2 is true'), e.g. *even though*, Fr. *quand même*, etc.
- expressions of **disjunction** ('whether ARG-1 is true or not, ARG-2 is true', e.g. Germ. *ob-wohl*, Lat. *sive*)
- expressions of **ineffectiveness** ('ARG-1 has no effect, ARG-2 holds', e.g. with expressions of contempt such as *spite*, e.g. Germ. *trotz*,)
- emphatic expressions of **simultaneity** ('while ARG-1 is true, ARG-2 is also true', e.g. Engl. *while*)
- expressions of **admission** ('let ARG-1 be the case, ARG-2 still holds')
- expressions of **irrelevance** ('ARG-1 does not matter, ARG-2 holds')

Volker Gast   Concessives and corpus-based typology

Some examples from Tzotzil

▶ Expressions of permission and irrelevance as markers of concessivity seem to be widespread in the languages of the world.

(23) **Ak'o mi** chahatav-e, chasmaj ?onox li Xun-e.
let.IMP Q you.flee-CL he.will.hit.you TEMP DET Xun-CL
'Although you flee, John will hit you.'

(24) **Manchuk** lubemon, chibat.
CANCEL I.am.tired I.will.go
'I'll go although I am tired / I would go if I weren't tired.'

(25) **Manchuk** li vo'-e, lek.
CANCEL DET rain-CL good
'It would be good if it weren't raining.'
(Haviland 1981)

A closer look at Tzotzil

- ▶ Tzotzil has a rich system of subordinators, many of them based on verbal or adjectival predicates.
- ▶ Specific subordinators can be combined, expressing multiple relations holding between propositions (e.g. conditional and temporal).

(26) **K'alal mi** chcham chkiltik ti   htotik-e,   sk'an htihtik
      when if   dies   we.see DET our.father-CL, NEC we.play
      ti   hbintike.
      DET our.pots
      'If our father (the sun) dies (solar eclipse), we have to make music (when he dies).' Gast (1998)

Temporal relations with attitudinal concessivity

- ▶ Tzotzil has a suffix *-uk* (also used as a subjunctive marker) that expresses what I call 'attitudinal concessivity' when suffixed to adverbial subordinators.

- ▶ If the temporal subordinator *k'alal* takes *-uk*, it expresses a negative attitude towards the simultaneous occurrence of two events.

(27) **K'alal** chiveʔ-e, chinop.
      when I.eat-CL I.fill.up
      'When I eat, I satiate my hunger.'

(28) **K'alal-uk** chiveʔ-e, lah htiʔ kok'.
      when-SUBJ I.eat-CL CP I.bite my.tongue
      'When I ate, I bit my tongue.'
      Gast (1998)

Purposive relations with attitudinal concessivity

- ▶ When the purposive subordinator *sventa* takes *-uk*, it implies that the main clause event happened despite (subjectively) unfavourable circumstances.

(29) Chibat   ta Chamo? **sventa**   hchan bats'i k'op
     I.will.go to Chamula in.order.to I.learn real   language
     'I'll go to San Juan Chamula in order to learn Tzotzil.'

(30) Chibat   ta Chamo? **sventa-uk**   hchan to   jutuk.
     I.will.go to Chamula in.order.to-SUBJ I.learn still a.bit
     'I will go to San Juan Chamula (despite unfavourable circumstances) in order to learn a bit (of Tzotzil).'
     Gast (1998)

Attitudinal concessivity and epistemic concessives

▶ The concessive subordinators *manchuk* and *k'u chaʔal* can also take *-uk*.

(31) **K'u-uk**   **chaʔal**   mi yanihemik xa   hlomuk ti
     how-SUBJ CONC-SUBJ Q get.lost   already some   DET
     k'opetik-e,   ...
     our.ideas-CL
     'Although some of our ideas (cultural values) have been lost, ...'
     Gast (1998: 100)

(32) **K'u-uk**   **chaʔal-uk**   mi yantik   xhel   ti
     how-SUBJ CONC-SUBJ Q piece.by.piece changes DET
     hvoʔne hkostumpretik-e, ...
     our.traditions-CL
     'Even though our traditions are gradually changing, ...'
     Gast (1998: 104)

Concessive subordinators in English, Spanish and German: A corpus-based study

- Pilot study with 150 richlyl annotated examples from the Europarl corpus (English originals and their Spanish and German translations, only speeches from 1999).
- Questions:
    - How do the semantic/pragmatic variables relate to each other?
    - What semantic/pragmatic factors determine the distribution of the various markers of concessivity, and in what way?
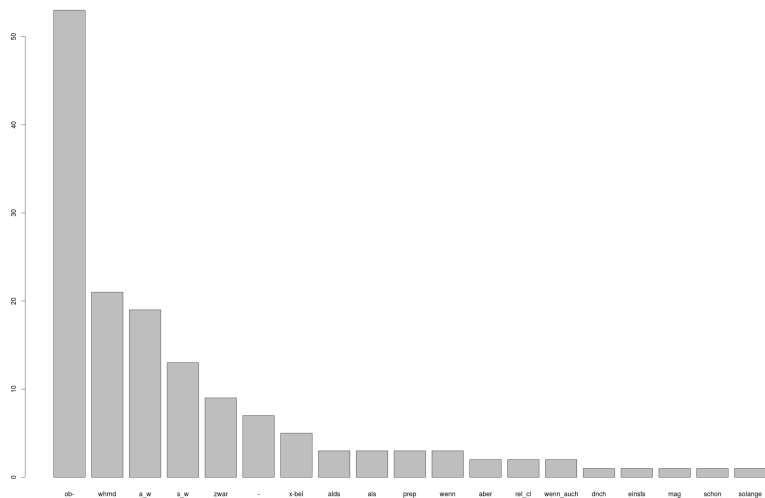
An annotation scheme

- ► Concessive predications are separated into two arguments, ARG-1 and ARG-2 (main assertion).
- ► Focus on three variables:
  - ► encyclopaedic background assumptions:
    con, *pro*
  - ► status of ARG-1:
    hypothetical, *presupposed*, *asserted*
  - ► paradigmatic relationship between arguments:
    vertical, *horizontal*, *none*
- ► More variables that can be extracted automatically:
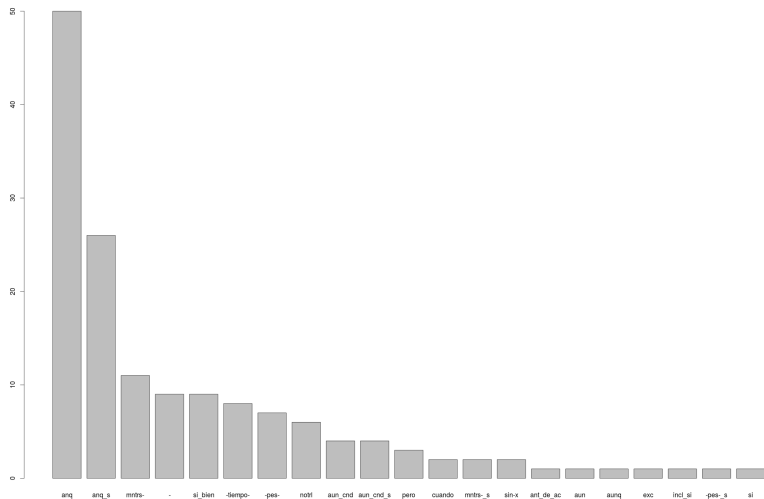 - ► relative ordering of the arguments
  - ► length of the arguments

Volker Gast Concessives and corpus-based typology

English strategies and their distribution

Volker Gast Concessives and corpus-based typology
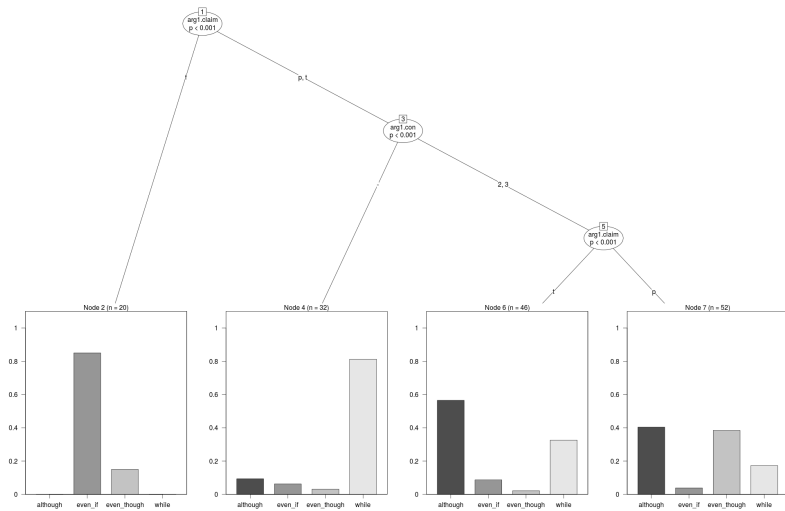
German translations

Spanish translations

Volker Gast Concessives and corpus-based typology
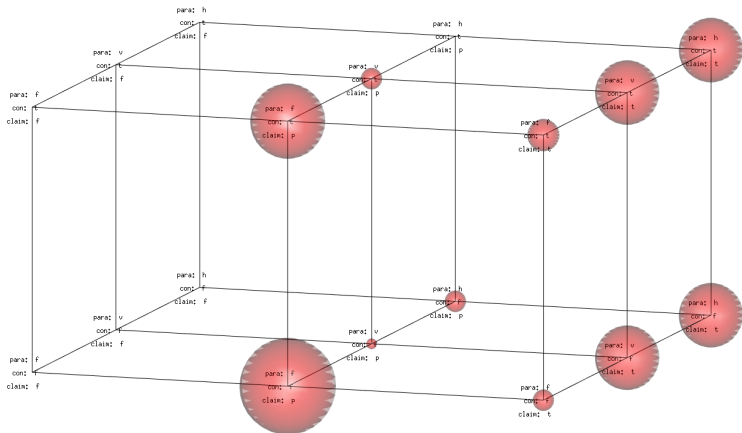
Towards a multidimensional representation: *although*
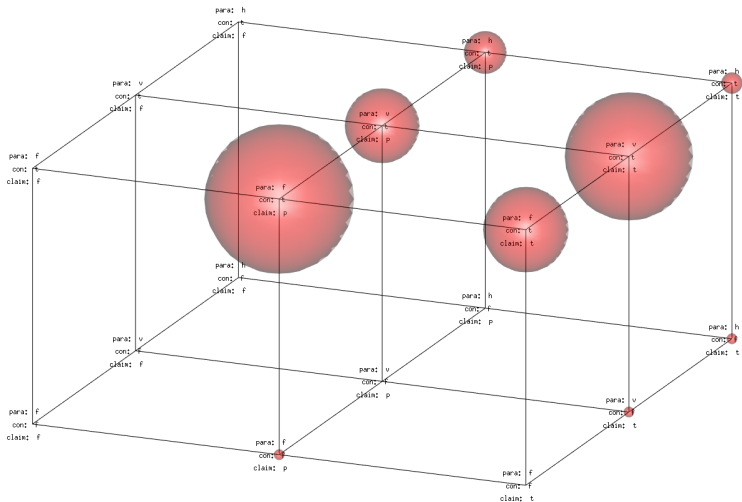
Towards a multidimensional representation: *while*

Volker Gast   Concessives and corpus-based typology
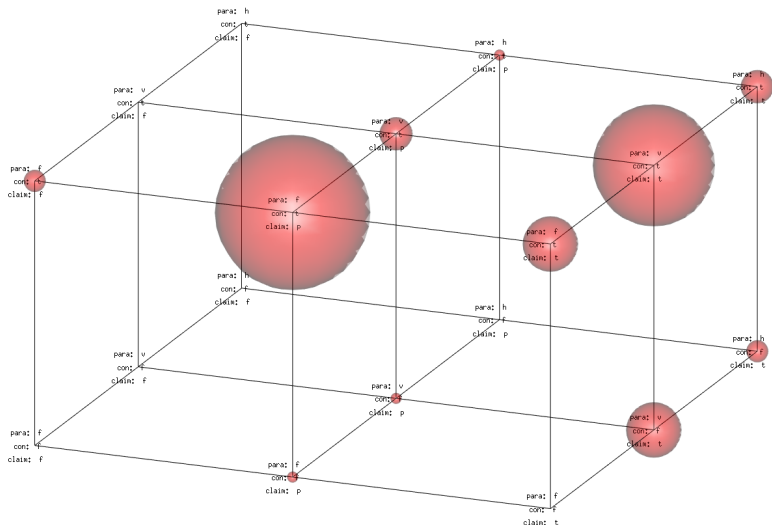
Towards a multidimensional representation: *obwohl*
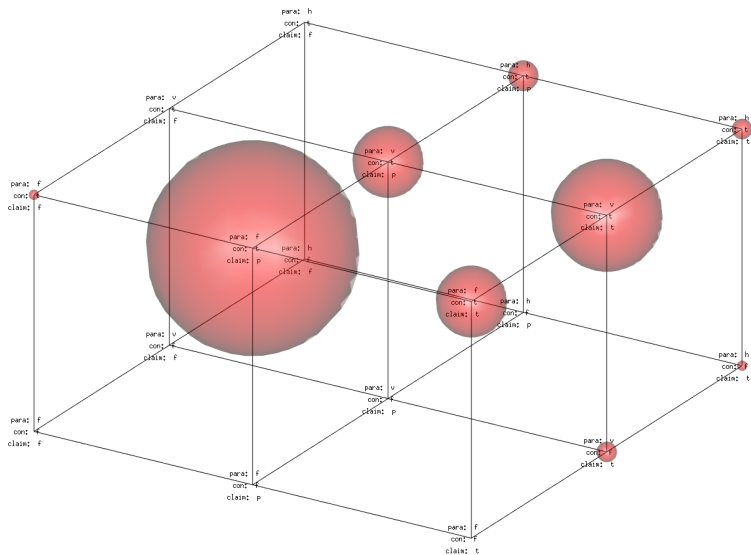
Volker Gast   Concessives and corpus-based typology

Towards a multidimensional representation: *aunque*/ind

Volker Gast   Concessives and corpus-based typology

Towards a multidimensional representation: *aunque*/subj
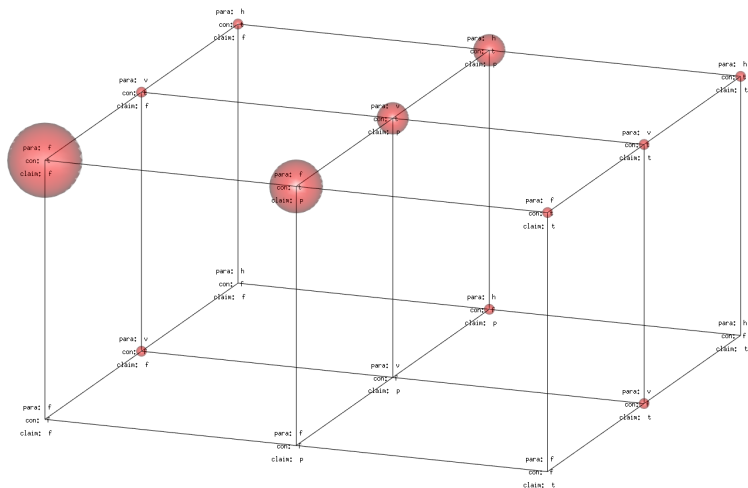
... Thanks for your attention!