

Towards a corpus-based typology of clause linkage

A case study of cross-clausal extraction

Volker Gast*, Balthasar Bickel⁺, Lennart Bierkandt*

*Friedrich-Schiller-Universität Jena, ⁺Universität Zürich

05.09.2011

Clause linkage

- (1) [Bill ordered a beer] and [Mary ordered a whiskey]. (coordination)
- (2) [I don't think [that he will ever change]].
(finite complement clause)
- (3) [I asked him [to slow down]]. (control infinitive)
- (4) [He risked [getting caught]]. (gerund)
- (5) [[He went to the library] [to lend a book]]. (adverbial infinitive)
- (6) [[Walking home], he met his brother-in-law]. (adjunct participle)
- (7) [[With John driving], there was no need to worry].
(absolute participle)
- (8) [Bill approached the man [who was drinking a Martini]].
(relative clause)

The typology of clause linkage

- Major parameters of variation:
 - Properties of the attachment site (e.g. verb, noun, sentence)
 - Relation of dependency (adjunct, complement, no dependency)

	nominal projection	verbal projection
adjunction	relative clause	adverbial clause
complementation	nominal complement clause	verbal complement clause

(9) The man [who lives next door] . . .

(10) It was raining [when I came home].

(11) The claim [that grammar is innate] was first made . . .

(12) Jack believes [that grammar is innate].

Some refinements

- Lehmann (1988), Bickel (1991): towards a multiparametric typology of clause linkage
- Lehmann (1988):
 - **Autonomy** vs. **integration**
(hierarchical downgrading, syntactic level)
 - **Expansion** vs. **reduction**
(desententialization of subordinate clause, grammaticalization of main verb)
 - **Isolation** vs. **linkage**
(interlacing, explicitness of linking)

Autonomy vs. integration

- Hierarchical downgrading: parataxis ↔ embedding
- Syntactic level: sentence ↔ lexical item

(13) [_S London is the capital of England] and
[_S Paris is the capital of France].

(14) ye nam su-ab is-omei (Usan)
I tree cut-SS go.down-1SG.PAST
'I cut the tree and went down.' (Fedden forthcoming)

(15) [_S When he found out that grammar was not innate],
[_S Jack was very disappointed].

(16) He [_{VP} claimed_V [that grammar is innate]].

(17) wa-rim-ak-ni-n-m (Alamblak)
IMP-DIR-get-go-2SG.A-3PI.U
'Get them and go away from me.'
(Foley & Van Valin 1984: 262)

Expansion vs. reduction

- Desentailization of subordinate clause (traditionally finiteness)
- Grammaticalization of main verb

(18) He risked [getting caught].

(19) Ch-ba ve'-ik-on. (Tzotzil)
ICP.IND-go eat-SBJ.1ABS
'I'll go and eat.'

Isolation vs. linkage

- Explicitness of linking ([a]syndesis)
- Interlacing (\rightarrow cross-clausal dependencies)

(20) John entered the room, Mary left.

(21) Because/when/although John entered the room, Mary left.

Cross-clausal dependencies

- Cross-clausal dependencies are not just symptoms of clause linkage; they actively participate in the establishment of inter-clausal relations.
- Example: Purpose clauses are often (asyndetically) juxtaposed to the main clause.
- The purposive interpretation is closely associated with subject identity and 'non-autonomous' TAM-marking (e.g. posterior, non-realis).

(22) ch-i-muy h-tuch' i tahchuh-e (Tzotzil)
 ICP-1ABS-climb 1ERG-cut DET lentinus.mushroom
 'I climbed up to pick the lentinus mushroom.'
 (Haviland 1993: 35)

Three major types of cross-clausal dependencies

- Argument-related dependencies
(e.g. argument sharing, argument trespassing)
- Predicate-related dependencies (e.g. TAM)
- Scope dependencies
(e.g. focus particles, illocutionary operators)

Argument-related dependencies

- Raising:
'Mismatch' between semantic and syntactic arguments
- Control:
One constituent corresponds to two (semantic) arguments

(23) I want you to do this immediately.
 ← ARG_{syn} ARG_{sem} →

(24) I asked him to pass me a beer.
 ← ARG_{syn/sem} ARG_{sem} →

More types of argument-related dependencies

- Argument trespassing: prolepsis
- Agreement across clause boundaries ('external agreement', cf. Creissels forthcoming)

(25) Vides-ne me ut rapior? (Latin)
 you.see-Q me that I.am.kidnapped
 'Don't you see (me) that I am being kidnapped?'

(26) mōla rasadi w-uL'-ī [š^wela-l̄-a m-āne
 Molla Rasadi M-die-HPL graveyard-OR-ALL HPL-go.PROG
 bak'-ī goli]
 HPL-be-HPL COP.HPL
 'Molla Rasadid died, and they were going to the graveyard.' (Creissels forthcoming)

The emergence of cross-clausal agreement

- Causative clauses in Tzotzil: grammaticalization of lexical verb

(27) Laj k-ak'-b-ot ve'-ik-ot
CP 1ERG-give/let-IO-2ABS eat-SBJ-2ABS
'I let you eat.'

Predicate-related dependencies

- *Consecutio temporum*: the tense of the subordinate clause is chosen in accordance with the tense of the main clause (and relative temporal order)
- Mood selection: the mood of the embedded clause depends on the matrix predicate

(28) Non dubito/dubitabam, quin hoc verum sit/esset. (Latin)

NEG I.doubt/doubted COMP this true is/was.SBJ

'I do/did not doubt that this is/was true.'

(29) Quiero que vayas/*vas (Spanish)

I.want that you.go.SBJ/you.go.IND

'I want you to go.'

Scope-dependencies

- In many languages, propositional operators (e.g. illocutionary force, tense) may have variable scope

- (30) ko-saja hand-a-ciy-e? (Chintang)
roam-CVB chat-PST-by.oneself-PST
'Did she chat by herself whilst roaming about?' or
'Did she roam about while chatting to herself?'
[CLLDCh 1R03S04.021]
- (31) Did he die singing?
a. Was he singing when he died?
b. Did he die while he was singing?
c. Did he die and sing?

Cross-clausal dependencies in a typology of clause linkage

- Cross-clausal dependencies are partially grammatical and partially probabilistic
 - They need to be investigated on the basis of corpus data
- Project: A corpus-based typology of clause linkage
- Requirements: Richly annotated corpora
- Languages: English, Latin; Chintang, Nepali; Tzotzil; Nen-Tonda (PNG)
- Agenda
 - develop a system of variables;
 - annotate corpora, refine the system of variables;
 - extract language-specific generalizations from corpora;
 - extract typological generalizations from corpora.

Dimensions of typologizing

- Holistic vs. **parametric**
- Aprioristic vs. **emergent**
- Theory-driven vs. **data-driven**
- Categorical vs. **probabilistic**

Cross-clausal extraction and clause linkage

- Long-distance extraction as an instance of ‘argument trespassing’
- Filler-gap dependency establishes a syntactic link between the sentences

(32) What_i do you think [he will say ____i]?

Grammatical constraints

- Extraction is generally not possible from specific positions (e.g. determiners).
- Pied-piping is required in such cases.
- PPs: Pied-piping seems to be obligatory in some cases.

(33) *Whose_i do you think [_S[t_i son] will win the contest]?

(34) [Whose son]_i do you think [_St_i will win the contest]?

(35) ... [in what way]_i can they say [that they remain Anglican t_i]?

Semantic constraints

- Long-distance extraction is (generally assumed) not to be possible across factive predicates.

(36) What do you think [_She did]?

(37) *What do you know [_S(that) he did]?

(38) *What do you regret [_S(that) you did]?

Some questions

- What biases are there in the distribution of long-distance extraction relative to specific context features.
- Analysis of (explanation for) the facticity constraint?

A pilot study

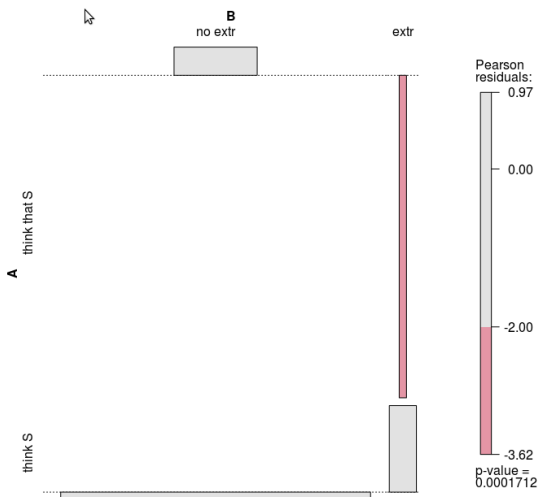
- Extraction of 400 examples from the BNC
- Coding for variables that are known/assumed to be sensitive to clause linkage
 - matrix predicates
 - presence of *that*
 - properties of extracted element (syntactic function and category)
 - properties of matrix and embedded subject (animacy, person)
 - properties of verbs (TAM)
- Quantitative analysis of data

Frequency of subordinating *that*

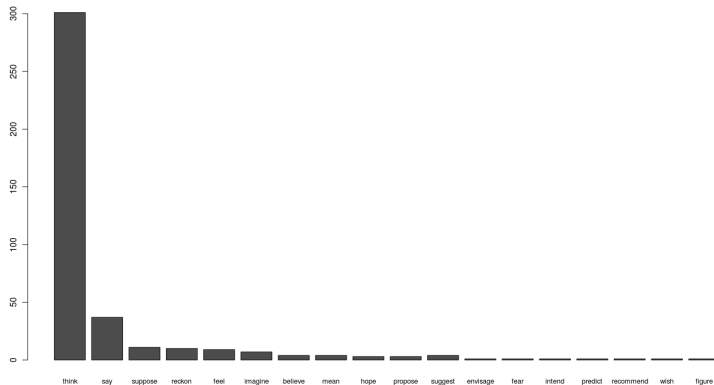
- Assumption: If cross-clausal extraction contributes to the degree of 'interlacing' (Lehmann 1988) between the clauses, *that* (as an overt marker of subordination) is expected to be less frequent in the relevant cases.
- Result: Subordinating *that* is underrepresented in clauses exhibiting cross-clausal extraction.

	think that S	think S	ratio
no extr	2665	36767	$\approx 1 \div 14$
extr	4	297	$\approx 1 \div 74$

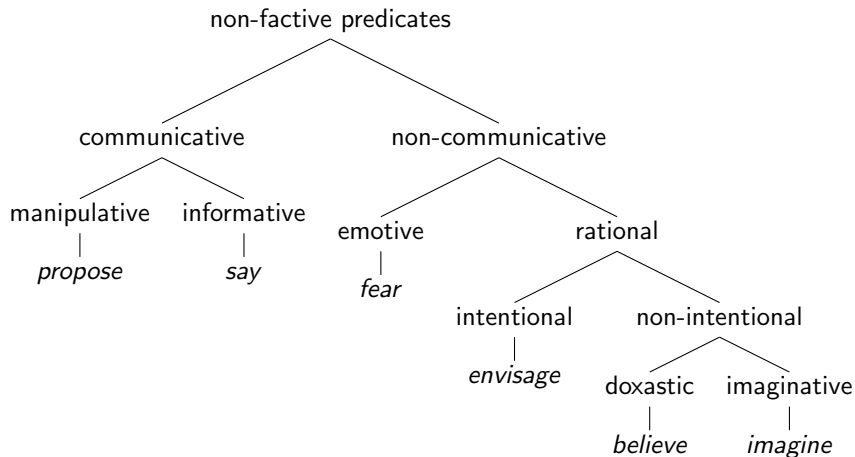
Frequency of subordinating *that*



Matrix predicates



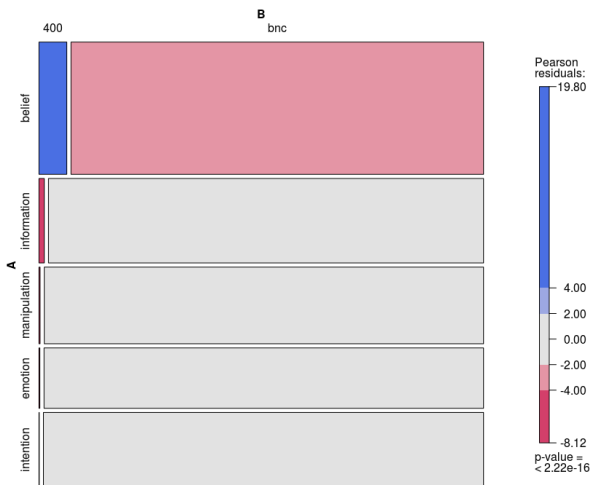
A taxonomy of (non-factive) predicates



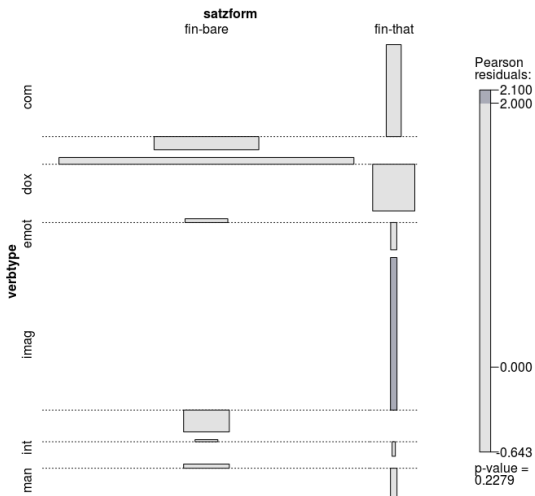
Some examples

- (39) manipulative
How do you propose this information be used?
- (40) informative
Where did you say you were going?
- (41) emotive
What do you fear it might be?
- (42) intentional
What did the participants intend should come about as a result?
- (43) doxastic
What do you think would happen to it?
- (44) imaginative
Where do you imagine we collect our stores if not in the city?

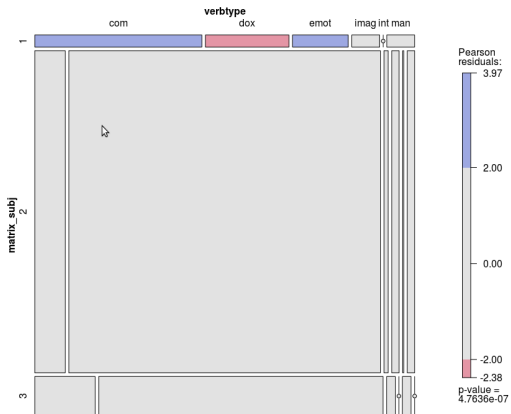
Frequencies of verb types



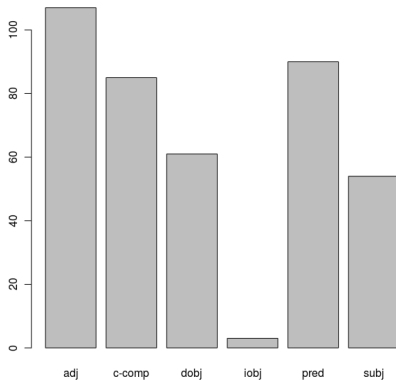
Verb type and *that*



Verb type and matrix subject

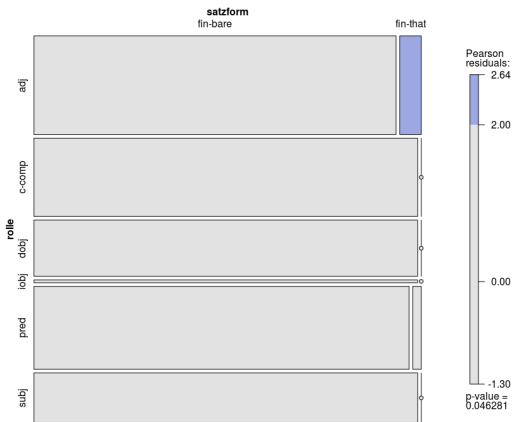


Syntactic role of extracted element



Syntactic role and use of *that*

- Relatively even distribution; overrepresentation of *that* with adjuncts



Syntactic role and *that*

- Why is *that* overrepresented with adverbials?
- Hypothesis: Adverbials do not represent a good instance of 'interlacing', as they could belong to either sentence.

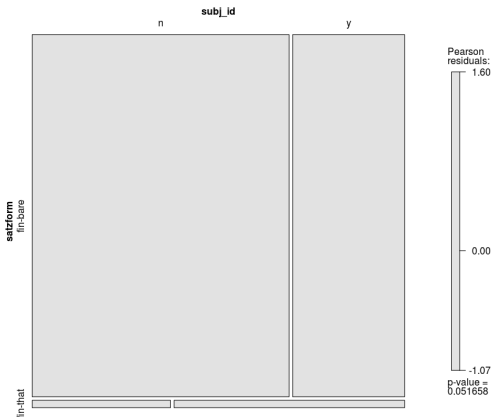
(45) ... why do you think that neither of them can see?

(46) When do you think that that erm stopped being popular?

(47) When does he believe that action will be taken?

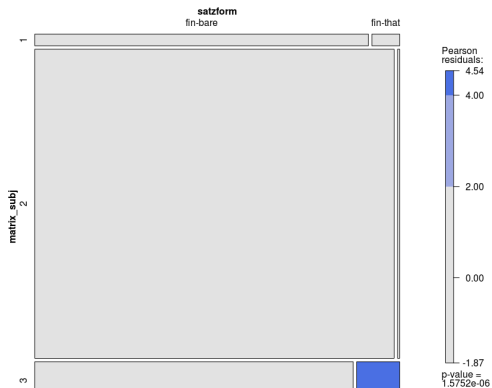
The factor of subject identity

- Subject identity as interlacing?



The category of person and *that*

- *That* is overrepresented with third person matrix subjects



Second person vs. the rest?

- Second person subjects often appear in quasi-parenthetical uses

(48) What do you think will he do?

(49) Was, glaubst du, wird er tun?

Other variables

- Animacy does not show any (interesting) correlations.
- TAM-related categories do not seem to have any major impact.
- Medium (spoken, written) seems to be a minor factor.

The semantics of cross-clausal extraction

- Wh-questions as instructions to select from a set of propositions.
- They introduce an existential presupposition
- Propositions differ with respect to the referent corresponding to the wh-pronoun.

(50) Who ate the cake?

(51) $\exists x[x \text{ ate the cake}]$

(52) Select the correct proposition from the set
{Ernie ate the cake, Bert ate the cake}!

Wh-questions and presuppositions

- Presuppositions survive within the scope of a Wh-operator
- The presuppositions are shared by all members of the answer set.

(53) When did you stop smoking?

(54) {I stopped smoking a month ago, I stopped smoking a week ago ... }
→ I used to smoke.

Factive verbs in wh-questions

- Factive verbs introduce a factive presupposition (shared by all members of the answer set).
- *Know*: Each member of the answer set introduces a different presupposition, and all the presuppositions survive.

- (55) Who do you think killed Kennedy?
{I think that Bert killed Kennedy, I think that Ernie killed Kennedy}
no presupposition
- (56) *Who do you know (that) killed Kennedy?
{I know that Bert killed Kennedy, I know that Ernie killed Kennedy}
presuppositions: {Bert killed Kennedy, Ernie killed Kennedy}

Long extraction and the scope of a Wh-operator

- Long extraction with *know* (and other factive verbs): The inner proposition is not under discussion; it is presupposed.
- This seems to be disallowed in English!
- Descriptive generalization: English Wh-questions are always questions about the 'local proposition' (*Who killed Kennedy?*).
- The matrix predicate merely provides a frame of evaluation (a doxastic system).

Factive vs. non-factive predicates: an example

- Problem with factive predicates: When interpreted with narrow scope, all options available to the addressee are presupposed!

- (57) a. Who do you think killed Kennedy?
b. In your doxastic system: Who killed Kennedy?
c. Select the correct proposition from among the proposition(s) that you consider likely: {Bert killed Kennedy, Ernie killed Kennedy}
- (58) a. *Who do you know (that) killed Kennedy?
b. In your doxastix system: Who killed Kennedy?
c. Select the correct proposition from among the propositions that you consider likely: {Bert killed Kennedy, Ernie killed Kennedy} (both presupposed!)

Summary

- Quantitative approach to typology
- Investigation of clause linkage on the basis of richly annotated corpora
- Case study: Non-local dependencies in English
- Outlook: generalizations made on the basis of crosslinguistic corpora
- Major challenge: How to annotate corpora for fine-grained variables (e.g. operator scope).