



The order of nominal subject and pronominal object in the *Mittelfeld*:

A case of free syntactic variation?

Kris Heylen



University of Leuven
RU Quantitative Lexicology and Variational Linguistics

Purpose of the talk

Type of word order variation in the “Mittelfeld” with two (nearly) always exchangeable variants:

1. “Free variation” or constraints?
 - Both language internal and external factors with gradient effects
2. How can such a subtle variation be studied empirically?
 - Statistically validated corpus analysis
3. How can it be modelled theoretically?
 - Usage-based Construction Grammar(s)

Overview

1. Situating word order variation in the “Mittelfeld”
 - The order of nominal subject and pronominal object
 - in different varieties of German
2. Corpus analysis
 - Data collection & annotation
 - The effect of individual factors
 - The combined effect of factors
3. Theoretical Interpretation
 - Variation in a usage-based Construction Grammar

Word order variation in German

- Relative order of bare-NP verb arguments in Mittelfeld: subject (N), accusative object (A), dative object (D)
- Both in main and subordinate clauses; with nominally and pronominally realized NP's

	Vorfeld	Kl. 1	Mittelfeld		Klammer 2	Nachfeld
$N_n A_n$	Dort	hat	der Lehrer	den Schüler	gesehen	als...
$A_n N_n$	“	“	den Schüler	der Lehrer	“	“
$D_p N_n$		dass	ihr	die Mutter	geholfen hat	
$N_n D_p$		“	die Mutter	ihr	“	

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
N_n A_n	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
N_n D_n	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
D_n A_n	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
N_p A_p	Dann hab ich ihn gesehen	*
N_p D_p	Dann bin ich ihr gefolgt	*
A_p D_p	Sie hat es ihm gegeben	??
N_p A_n	Dann hat er den Mann gesehen	*
N_n A_p	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
N_p D_n	Dann ist er der Frau gefolgt	*
N_n D_p	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
A_p D_n	Sie hat es dem Mann gegeben	??
A_n D_p	Sie hat ihm das Buch gegeben	??

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
N_n A_n	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
N_n D_n	Dann ist der Mann gefolgt	Dann ist der Mann gefolgt
D_n A_n	Sie hat der Frau das Buch gegeben	Sie hat das Buch der Frau gegeben
N_p A_p	Dann hab ich ihn gesehen	*
N_p D_p	Dann bin ich ihr gefolgt	*
A_p D_p	Sie hat es ihm gegeben	??
N_p A_n	Dann hat er den Mann gesehen	*
N_n A_p	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
N_p D_n	Dann ist er der Frau gefolgt	*
N_n D_p	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
A_p D_n	Sie hat es dem Mann gegeben	??
A_n D_p	Sie hat ihm das Buch gegeben	??

Widely discussed in the literature:
more or less clear constraints

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
$N_n A_n$	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
$N_n D_n$	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
$D_n A_n$	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
$N_p A_p$	Dann hab ich ihn gesehen	*
$N_p D_p$	Dann bin ich ihr gefolgt	*
$A_p D_p$	Sie hat es ihm gegeben	??
$N_p A_n$	Dann hat er den Mann gesehen	*
$N_n A_p$	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
$N_p D_n$	Dann ist er der Frau gefolgt	*
$N_n D_p$	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
$A_p D_n$	Sie hat es dem Mann gegeben	??
$A_n D_p$	Sie hat ihm das Buch gegeben	??

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
$N_n A_n$	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
$N_n D_n$	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
$D_n A_n$	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
$N_p A_p$	Dann hab ich ihn gesehen	*
$N_p D_p$	Dann bin ich ihr gefolgt	*
$A_p D_p$	Sie hat es ihm gegeben	??
$N_p A_n$	Dann hat er den Mann gesehen	*
$N_n A_p$	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
$N_p D_n$	Dann ist er der Frau gefolgt	*
$N_n D_p$	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
$A_p D_n$	Sie hat es dem Mann gegeben	??
$A_n D_p$	Sie hat ihm das Buch gegeben	??

No variation

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
N _n A _n	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
N _n D _n	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
D _n A _n	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
N _p A _p	Dann hab ich ihn gesehen	*
N _p D _p	Dann bin ich ihr g	*
A _p D _p	Sie hat es ihm gegeben	??
N _p A _n	Dann hat er den Mann gesehen	*
N _n A _p	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
N _p D _n	Dann ist er der Frau gefolgt	*
N _n D _p	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
A _p D _n	Sie hat es dem Mann gegeben	??
A _n D _p	Sie hat ihm das Buch gegeben	??

No variation

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
N _n A _n	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
N _n D _n	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
D _n A _n	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
N _p A _p	Dann hab ich ihn gesehen	*
N _p D _p	Dann bin ich ihr gefolgt	*
A _p D _p	Sie hat es ihm gegeben	??
N _p A _n	Dann hat er den Mann gesehen	*
N _n A _p	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
N _p D _n	Dann ist er der Frau gefolgt	*
N _n D _p	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
A _p D _n	Sie hat es dem Mann gegeben	??
A _n D _p	Sie hat ihm das Buch gegeben	??

Combinations of 2 verb arguments, nominal or pronominal:

	Variant 1	Variant 2
N _n A _n	Dann hat der Mann den Hund gesehen	Dann hat den Hund der Mann gesehen
N _n D _n	Dann ist der Mann dem Hund gefolgt	Dann ist dem Hund der Mann gefolgt
D _n A _n	Sie hat der Frau das buch gegeben	Sie hat das Buch der Frau gegeben
N _p A _p	Dann hab ich ihn gesehen	*
N _p D _p	Dann bin ich ihr gefolgt	*
A _p D _p	Sie hat es ihm gegeben	??
N _p A _n	Dann hat er den Mann gesehen	*
N _n A _p	Dann hat der Mann ihn gesehen	Dann hat ihn der Mann gesehen
N _p D _n	Dann ist er	
N _n D _p	Dann ist die Frau ihm gefolgt	Dann ist ihm die Frau gefolgt
A _p D _n	Sie hat es dem Mann gegeben	??
A _n D _p	Sie hat ihm das Buch gegeben	??

Variation but no clear constraints

Word order variation in German

Nominal subject and pronominal object in the Mittelfeld

		Mittelfeld		
Object-first	Später nahm	ihn	der Chef	beiseite
Subject-first	...wie	viele Frauen	sich	wünschen

examples from Negra corpus

- In both main and subordinate clauses, with personal and reflexive pronouns, accusative and dative objects
- Object-first is most common, but there are no clear differences in meaning or grammaticality

⇒ **FREE VARIATION??**

Word order variation in German

- Variation is well known in reference grammars, and fairly common but little studied in the word order literature:
- Generative paradigm
 - Lenerz (1994): optional movement of S_{nom} to SpecIP
 - Müller (2002): optional S-raising to SpecT (matched constraints)
 - ⇒ identify no motivation/trigger for movement
- Functional paradigm
 - Hoberg (1997): marked subject-first seems to occur more often with proper noun subjects, but no clear tendencies

⇒ Free variation / No constraints ??

Word order variation in German

main shortcomings of previous research:

- Language external factors were not considered:
 - possible differences in word order preferences between varieties

THIS STUDY: - regional varieties of German (A, D, CH)
- spoken vs. written language use

- Methodological deficiencies:
 - German word order studies haven been mainly based on introspective grammaticality judgements that are notoriously unreliable when factors with gradient effects are involved

THIS STUDY: statistically validated corpus analysis

Corpus analysis: Data

- Written Data: Newspaper corpora, 1 per country
 - Germany (D): *Frankfurter Rundschau*⁺ 995 obs.
 - Austria (A): *Salzburger Nachrichten*^{*} 651 obs.
 - Switzerland (CH): *St.-Galler Tagblatt*^{*} 799 obs.
 - Spoken Data: German radio talkshows[°] 438 obs.
-
- 2883 obs.

+ NEGRA corpus, compiled at *University of Saarbrücken*

* COSMAS corpora compiled at *Institut für deutsche Sprache*

° IDS-Deutsches Spracharchiv: Dialogstrukturenkorpus

Corpus analysis: Data

Steps in Corpus analysis:

- Definition of a relevant observation: “Mittelfeld” contains a nominally realized subject and one pronominally realized accusative or dative object
- Observations retrieved manually (A ,CH, spoken) and semi-automatically (D)
- Coded for language internal and external factors
- Resulting dataset analyzed statistically

Corpus analysis: Data

Observations stratified for 2 language-external factors:

1. Country: D, A, CH
2. Medium: written / spoken

Coded additionally for 5 language-internal factors:

1. Subject noun type: proper >< common (cf. Hoberg 1997)
2. Subject animacy: animate >< inanimate (cf. Hoberg 1997)
3. Subject length: number of letters (cf. Behaghel 1932)
4. Object pronoun type: personal >< reflexive (cf. Lenerz 1977)
5. Clause type: main >< subordinate

Corpus analysis: overview statistics

Variation in the varieties

- Object-first clearly default, in all varieties
- Marked subject-first order significantly more common in spoken than in written
- Marked subject-first order significantly more common in D than in A and CH, but no significant difference between A and CH

	Subj.-first	Object-first
spoken	153 (35%)	285 (65%)

written/D	106 (11%)	889 (89%)
-----------	-----------	-----------

A	88 (6%)	613 (94%)
CH	43 (5%)	756 (95%)

Statistics		
spoken~written	p < 0.01	OR 4,51
D ~ A	p < 0.01	OR 2,20
D ~ CH	p < 0.01	OR 2,21
A ~ CH	p > 0.1	



Corpus analysis: separate factors

1. Subject noun type

		Subject-first	Object-first	stats
sp.	Proper	19 (40%)	29 (60%)	p=0.06 or: 1.47
	Common	134 (35%)	256 (65%)	

wr. D	Proper	21 (15%)	115 (85%)	p<0.05 or: 1.68
	Common	85 (10%)	774 (90%)	

A	Proper	11 (11%)	88 (89%)	p<0.01 or: 2.48
	Common	27 (5%)	525 (95%)	
CH	Proper	5 (4%)	120 (96%)	p=0.30 or: 0.75
	Common	38 (6%)	636 (94%)	

- No overall effect (only D/written and A)
- Effect not significantly different in varieties

Corpus analysis: separate factors

2. Subject animacy

		Subject-first	Object-first	stats
sp.	Inanimate	47 (20%)	186 (80%)	p<0.01 or: 0.23
	Animate	106 (52%)	99 (48%)	

D	Inanimate	24 (6%)	357 (94%)	p<0.01 or: 0.43
	Animate	82 (13%)	532 (87%)	

A	Inanimate	5 (2%)	250 (98%)	p<0.01 or: 0.52
	Animate	33 (8%)	363 (92%)	
CH	Inanimate	12 (4%)	320 (96%)	p=0.06 or: 0.75
	Animate	31 (7%)	436 (93%)	

- Effect in all varieties (only marginal in CH)
- Effect not significantly different in countries
- Effect significantly stronger in spoken than written!! (BrD p<0.05)

Corpus analysis: separate factors

3. Subject length

		Subject-first	Object-first	stats
sp.	Short (≤ 10)	29 (23%)	97 (77%)	p=0.4
	Long (+10)	124 (43%)	166 (57%)	or: 0.45
D	Short (≤ 10)	48 (17%)	227 (83%)	p<0.01
	Long (+10)	58 (8%)	662 (92%)	or: 2.41
A	Short (≤ 10)	13 (8%)	160 (92%)	p=0.10
	Long (+10)	25 (5%)	453 (95%)	or: 1.50
CH	Short (≤ 10)	16 (9%)	171 (91%)	p<0.05
	Long (+10)	27 (4%)	585 (96%)	or: 2.05

- significant in written/D and CH, but not in spoken and A
- Effect not significantly different in varieties

Corpus analysis: separate factors

4. Object pronoun type

		Subject-first	Object-first	stats
sp.	Personal	45 (29%)	111 (71%)	p<0.04 or: 0.65
	Reflexive	108 (38%)	174 (62%)	

wr. /D	Personal	38 (21%)	141 (79%)	p<0.01 or: 2.97
	Reflexive	68 (8%)	748 (92%)	
A	Personal	15 (12%)	108 (88%)	p<0.01 or: 3.05
	Reflexive	23 (4%)	505 (96%)	
CH	Personal	10 (7%)	136 (93%)	p=0.38 or: 1.38
	Reflexive	33 (5%)	620 (95%)	

- Effect in spoken, written/D and A but not in CH
- Effect significantly different between spoken and written!!!

Corpus analysis: separate factors

5. Clause type

		Subject-first	Object-first	stats
sp.	subordinate	137 (51%)	130 (49%)	p<0.01 or: 10.02
	main	16 (9%)	155 (91%)	

wr. D	subordinate	78 (24%)	243 (76%)	p<0.01 or: 7.41
	main	28 (4%)	646 (96%)	
A	subordinate	34 (16%)	177 (84%)	p<0.01 or: 18.93
	main	4 (1%)	436 (99%)	
CH	subordinate	37 (18%)	166 (82%)	p<0.01 or: 21.91
	main	6 (1%)	590 (99%)	

- In all varieties marked order sig. more often in subordinates
- significantly stronger in A+ CH than in D (BrD $p < 0.05$)
- Strongest factor of all!!



Corpus analysis: combined effect

- Not enough to take into account separate effect, because in real usage, **all factors are at play simultaneously**. Questions:
 - Taken all the factors together, **which actually play a role**?
 - What is the **combined effect** of the factors?
 - What is **each factor's contribution** to the combined effect?
 - Which factor is **most important**?
- Logistic regression model (comparable to VARBRUL):
$$\text{Ln}(\text{odds}_{\text{subject first}}) = \text{intercept} + \text{weight1} * \text{factor1} + \text{weight2} * \text{factor2} + \dots$$

Corpus analysis: combined effect

step	Factor	weights spoken	weights wr./D	weights A+CH
1	CLAUSE TYPE (Subordinate >< main)	9.8	7.5	19.1
2	SUBJECT ANIMACY (animate >< inanimate)	4.2	4	1.7
3	PRONOUN TYPE (personal >< reflexive)	0.7	2.5	2.4
4	SUBJECT LENGTH (short >< long)	0.8	2	1.9

stepwise selection
on written/D data
(subj-first modelled)
same model on
other datasets for
comparison

- Clause type strongest effect in all varieties
- Effect of clause type sign. different in D >< A+CH
- Effect of pronoun type sign. different in spoken >< written

Corpus analysis: conclusion

Linguistic:

- No “free variation” but graded constraints
- Both language internal and language external
- Language internal constraints do not necessarily have the same effect in the different (external) varieties

Methodological:

- Multivariate statistical analysis of corpus data makes it possible to uncover graded constraints, even in cases with very subtle differences between variants

Theoretical Interpretation

- In principle, results of corpus analysis are theory independent and constitute an explanandum for every framework
- BUT: Which theoretical model is best suited to incorporate the graded effect of both internal and external factors observed in this variation?
- Here, I will argue for a usage-based Construction Grammar approach

Theoretical Interpretation: ub-CxG

- Family of Theories: Fillmore & Kay 1999, Goldberg 1995, Bybee & Hopper 2000, Tomasello 2003, Barlow & Kemmer 2000, Langacker 1999
- All linguistic items are considered as principally similar form-meaning pairings, i.e. constructions, organized in a network, the *constructicon*
- Here, focus on 2 properties of CxG:
 - Emergence from usage
 - network of constructions with varying degrees of schematicity

Theoretical Interpretation: ub-CxG

Emergence from usage:

- All constructions are acquired bottom-up from usage
- Schematic ones through pattern recognition
- Constructions become entrenched through frequent usage
- Accounts naturally for external variation:
 - Usage occurs within social communities and within each community different constructions will get more entrenched
 - Grammar will reflect community stratification, variation is inherent in this grammar model
 - Can also be applied to the regional and medium differences observed in our study

Theoretical Interpretation: ub-CxG

Network of constructions:

- Construction is a continuum with constructions ranging from fully lexical over idioms to highly schematic
- Inherent redundancy: mentally represented less schematic constructions can simultaneously instantiate more schematic constructions
- Word order can be fixed at different levels of schematicity
 - ⇒ strong effect of clause type seems to indicate that variation is better described at a lower level than the high level brace-construction



for further information:

<http://www.ling.arts.kuleuven.ac.be/qvl/>