



Descriptive, methodological and theoretical issues in syntactic variation research

Gert De Sutter

Dirk Speelman & Dirk Geeraerts



University of Leuven

RU Quantitative Lexicology and Variational Linguistics

General introduction

- Increased interest in syntactic variation research in recent years:
 - Dative alternation
 - Extraposition
 - Relative order of nominal/prepositional constituents

General introduction

- Increased interest in syntactic variation research in recent years:
 - Dative alternation
 - Extraposition
 - Relative order of nominal/prepositional constituents
- Yielding a better insight into some motivating principles of syntactic optionality:
 - Avoid ambiguity
 - Mark pragmatic value
 - Postpone complex constituents

General introduction

- But: most of the studies into syntactic variation address **interconstituent** variation, not **intraconstituent** variation:

General introduction

- But: most of the studies into syntactic variation address interconstituent variation, not intraconstituent variation:
 - The waiter brought **the wine we had ordered** to the table.
 - The waiter brought to the table **the wine we had ordered**.(HNPS, Arnold et al. 2000: 28)

General introduction

- But: most of the studies into syntactic variation address interconstituent variation, not intraconstituent variation:
 - The waiter brought the wine we had ordered to the table.
 - The waiter brought to the table the wine we had ordered.(HNPS, Arnold et al. 2000: 28)
- What about **intraconstituent variation**?
 - Less well-studied
 - More complex type of variation?
 - Other motivating principles?

Purpose

- Focus on a case of intraconstituent variation, viz. word order variation in Dutch clause final verb clusters:
 - Ik geloof niet dat Jan dat boek **gestolen**_{PART} **heeft**_{AUX}.
I believe not that Jan that book stolen has
 - Ik geloof niet dat Jan dat boek **heeft**_{AUX} **gestolen**_{PART}.
I believe not that Jan that book has stolen

Purpose

- Focus on a case of intraconstituent variation, viz. word order variation in Dutch clause final verb clusters
- Try to uncover the underlying **language-internal mechanisms** which govern the choice of word order
 - In actual language use
 - From a synchronic perspective
 - With considerable attention for methodological and analytical prerequisites of thorough empirical research
 - Using multivariate statistical techniques

Purpose

- Focus on a case of intraconstituent variation, viz. word order variation in Dutch clause final verb clusters
- Try to uncover the underlying language-internal mechanisms which govern the choice of word order
- Contribute to the recent trend towards **a fully-fledged theoretical and methodological framework** for usage-based grammatical variation research

TOC

- Presentation of the O.I.
- Corpus analysis
 - Design of the corpus study
 - List of explanatory variables
 - Multivariate statistical analysis
- Discussion
- Concluding remarks

Presentation of the O.I.



Word order variation in verb clusters

- Word order variation in Dutch clause final verb clusters
[...] dat Jan dat boek gestolen_{PART} heeft_{AUX}.
[...] dat Jan dat boek heeft_{AUX} gestolen_{PART}.
- No difference in grammaticality or meaning

Previous research

- Effect of several types of factors has been investigated since 1950's:
 - Prosodic variables
 - Morpho-syntactic variables
 - Semantic variables
 - Lectal variables
 - Discursive variables

Previous research

- Effect of several types of factors has been investigated since 1950's:
 - Prosodic variables
 - Morpho-syntactic variables
 - Semantic variables
 - Lectal variables
 - Discursive variables
- The choice for one of the syntactic variants is fundamentally free, but is not a case of free variation

Previous research

- However, previous analyses:
 - Different methods
 - Different empirical fundamentals
 - Unclear data selection
 - No / restricted use of statistical techniques

Previous research

- However, previous analyses:
 - Different methods
 - Different empirical fundamentals
 - Unclear data selection
 - No / restricted use of statistical techniques
- Hence, unanswered questions:
 - How do the already introduced variables relate to each other in their effect on the choice of word order?
 - What is the global effect and the predictive power of these variables?

Corpus analysis



Design of the study

- Corpora: ConDiv
 - Contains 47.397.581 tokens
 - No annotation
 - Stratified along a stylistic, diachronic and regional dimension

Design of the study

	Internet		Papers		
	IRC	Usenet	Popular papers		Quality papers
			Regional	National	
NL	6.965.291	7.748.436	<i>De Limburger</i> 1.680.636	<i>De Telegraaf</i> 1.590.581	<i>NRC - HB</i> 1.520.064
BE	8.207.007	4.980.780	<i>Het Belang van Limburg</i> 3.012.330 <i>De Gazet van Antwerpen</i> 3.068.405	<i>Het Laatste Nieuws</i> 3.486.911	<i>De Standaard</i> 3.228.910

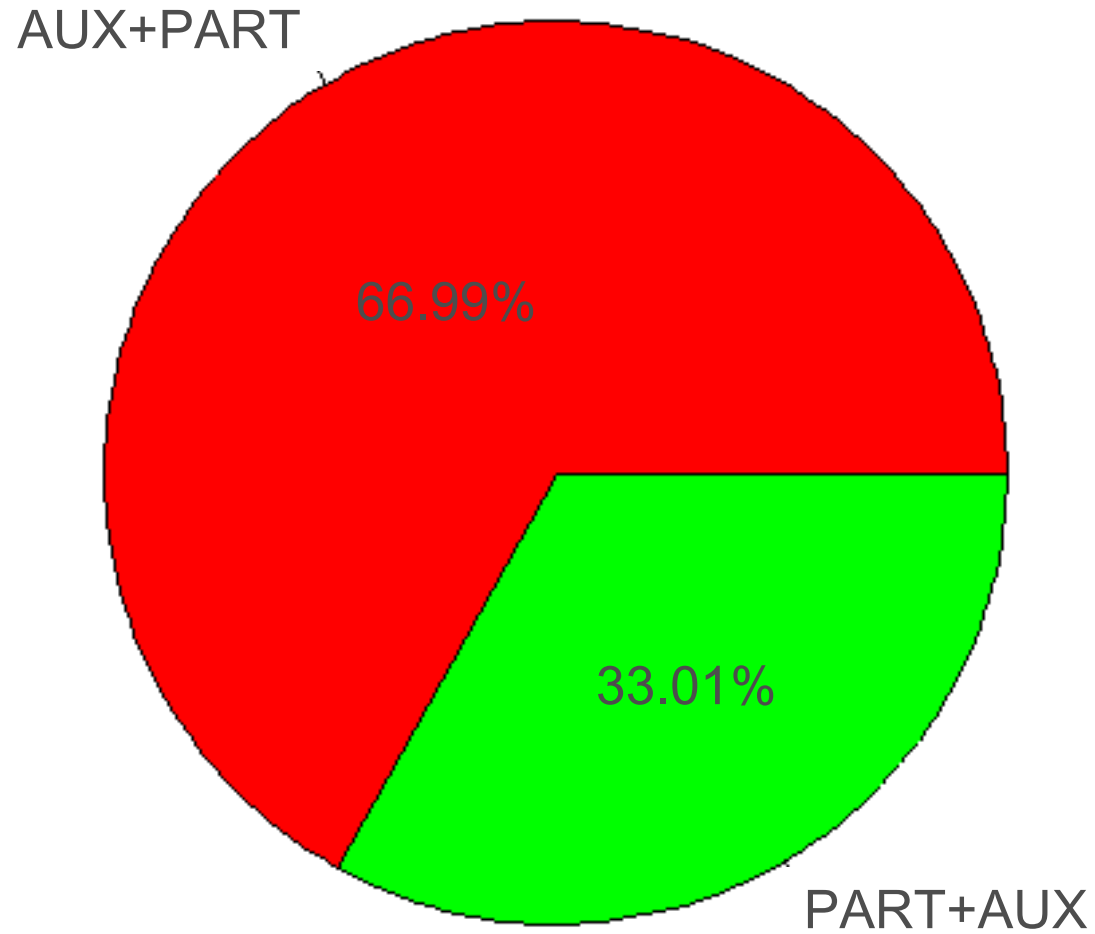
Design of the study

	Internet		Papers		
	IRC	Usenet	Popular papers		Quality papers
			Regional	National	
NL	6.965.291	7.748.436	<i>De Limburger</i> 1.680.636	<i>De Telegraaf</i> 1.590.581	<i>NRC - HB</i> 1.520.064
BE	8.207.007	4.980.780	<i>Het Belang van Limburg</i> 3.012.330 <i>De Gazet van Antwerpen</i> 3.068.405	<i>Het Laatste Nieuws</i> 3.486.911	<i>De Standaard</i> 3.228.910

Design of the study

- Data selection:
 - Bipartite verb clusters, consisting of a participle and the auxiliary verb *zijn*, *hebben* or *worden*
 - Complement clause introduced by *dat*
- Data extraction and processing: Abundantia Verborum
 - n = 2390
- Statistical analyses: R 2.0.1 (2003)

Design of the study



List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word
5. Inherence of the preceding constituent
6. Length of the middle field
7. Syntactic function of the extraposed constituent
8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word
5. Inherence of the preceding constituent
6. Length of the middle field
7. Syntactic function of the extraposed constituent
8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word
5. Inherence of the preceding constituent
6. Length of the middle field
7. Syntactic function of the extraposed constituent
8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle

Nominal variable:

→ Separable participle (e.g. oppgemerkt)

→ Non-separable participle (e.g. gemerkt)

6. Length of the middle field
7. Syntactic function of the extraposed constituent
8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle

Continuous variable:

→ Frequency of the lemma information in CELEX

3. Influence of the preceding constituent
6. Length of the middle field
7. Syntactic function of the extraposed constituent
8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. **Semantic-categorial status of the auxiliary / participle**

Nominal variable (Participle Disambiguating Algorithm):

→ Copulative use of *zijn* (PART = adjectival)

→ Passive use of *zijn* (PART = verbal)

→ Temporal use of *zijn* / *hebben* (PART = verbal)

→ Passive *worden* (PART = verbal)

→ Unclassifiable *zijn* (PART = ?)

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
- 4. Informationality of the preceding word**

Ordinal variable:

→ Highly informational (noun / verb / numeral)

→ Intermediately informational (adjective / adverb)

→ Low informational (pronouns / prepositions / conjunctions)

10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word
- 5. Inherence of the preceding constituent**

Nominal variable:

→ part of fixed verbal expression (significant association between participle and left collocate noun)

→ not part of fixed verbal expression (not significant association between participle and left collocate noun)

Ordinal variable (no. words between *dat* and verbal cluster):

→ 0-2 words

→ 3-5 words

→ 6-8 words

→ 9-11 words

→ 12-14 words

→ > 14 words

6. Length of the middle field

7. Syntactic function of the extraposed constituent

8. Accent distribution at the beginning of the verb cluster

9. Accent distribution at the end of the verb cluster

10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle

Nominal variable:

→ Adjunct / no extraposed constituent

→ Complement of V

→ Complement of preverbal N / A or Adv

7. Syntactic function of the extraposed constituent

8. Accent distribution at the beginning of the verb cluster

9. Accent distribution at the end of the verb cluster

10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word

Ordinal variable:

→ No. of unaccented syllables between preceding word accent and participial word accent

8. Accent distribution at the beginning of the verb cluster
9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word
5. Inherence of the preceding constituent

Ordinal variable:

→ No. of unaccented syllables between participial word accent and following word accent

9. Accent distribution at the end of the verb cluster
10. Syntactic persistence

List of explanatory variables

1. Morphology of the participle
2. Frequency of the participle
3. Semantic-categorial status of the auxiliary / participle
4. Informationality of the preceding word

Nominal variable:

- No preceding verb cluster
- Preceding [PART+AUX] cluster
- Preceding [AUX+PART] cluster

10. Syntactic persistence

Multivariate statistical analysis

- Statistical modeling of the variation:
 - Do all exp. variables have a significant effect on the choice of word order?
 - What is the relative impact of each exp. variable?
 - How do the different exp. variables relate to each other?
 - What is the collective effect of all exp. variables on the choice of word order?
 - What is the predictive power of the model?

Multivariate statistical analysis

- Binary logistic regression model:

$$\log \frac{p(Y = 1)}{p(Y = 0)} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

- Technical specifications:
 - Reference coding
 - No stepwise selection
 - Modeling odds in favour of [AUX+PART] order
 - Maximum Likelihood Estimation

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30		
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	<p>No association: O.R = 1 Positive association: O.R > 1 Negative association: O.R < 1</p>	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	The odds for [AUX+PART] word order increase 3.87 times when the participle is separable (vs. a non-separable participle)	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	<p>The odds for [AUX+PART] word order increase 1.41 times when the previous word is intermediately informational (vs. low informational)</p>	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	Significant positive association	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	Significant negative association	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	No significant association	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	Magnitude of the effect	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Morph_part: separ.	3.87	Synt_pers: [aux+part]	3.28
Informat.: intermed.	1.41	Synt_pers.: no cluster	1.72
Informat.: high	1.94	Frequency of participle	2.44 ^{-06*}
Fixed expr.: yes	2.26	Accent begin: 0 to 1 syll.	n.s.
middle field: 3-5 w.	2.03	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of V	0.47
middle field: >14 w.	1.98	extrapos: comp. of N	n.s.
Auxiliary: time	18.30	Magnitude of the effect	
Auxiliary: unclassif.	4.34		
Auxiliary: passive <i>zijn</i>	7.82		
Auxiliary: <i>worden</i>	11.73		

Predictor	O.R.	Predictor	O.R.
Auxiliary: time	18.30	Informat.: high	1.94
Auxiliary: <i>worden</i>	11.73	Synt_pers.: no cluster	1.72
Auxiliary: passive <i>zijn</i>	7.82	Informat.: intermed.	1.41
Auxiliary: unclassif.	4.34	Frequency of participle	2.44 ^{-06*}
Morph_part: separ.	3.87	Accent begin: 0 to 1 syll.	n.s.
Synt_pers: [aux+part]	3.28	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of N	n.s.
Fixed expr.: yes	2.26	Hierarchical ordering of the magnitude of the effects	
extrapos: comp. of V	0.47		
middle field: 3-5 w.	2.03		
middle field: >14 w.	1.98		

Predictor	O.R.	Predictor	O.R.
Auxiliary: time	18.30	Informat.: high	1.94
Auxiliary: <i>worden</i>	11.73	Synt_pers.: no cluster	1.72
Auxiliary: passive <i>zijn</i>	7.82	Informat.: intermed.	1.41
Auxiliary: unclassif.	4.34	Frequency of participle	2.44 ^{-06*}
Morph_part: separ.	3.87	Accent begin: 0 to 1 syll.	n.s.
Synt_pers: [aux+part]	3.28	Accent begin: 2 to 3 syll.	n.s.
middle field: 6-8 w.	2.29	Accent end: 0 to 1 syll.	n.s.
middle field: 9-11 w.	2.29	Accent end: 2 to 3 syll.	n.s.
middle field: 12-14 w.	2.57	extrapos: comp. of N	n.s.
Fixed expr.: yes	2.26	<p>Fit: $\chi^2 = 2480$, $df = 2367$, $p > .05$ Model LR = 789, $df = 22$, $p < .0001$ Prediction: $c = 0.8$ (100 bootstraps)</p>	
extrapos: comp. of V	0.47		
middle field: 3-5 w.	2.03		
middle field: >14 w.	1.98		

Variance Inflation Factors

Predictors	V.I.F.
Morphology of the participle: separable participle	1.843582
Informationality previous word: intermediate	3.330343
Informationality previous word: high	3.539587
Inherence previous constituent: fixed verbal expression	1.151188
Length of the middle field: 3 to 5 words	1.918343
Length of the middle field: 6 to 8 words	1.805834
Length of the middle field: 9 to 11 words	1.448133
Length of the middle field: 12 to 14 words	1.187800
Length of the middle field: >14 words	1.122027
Status of the participle / auxiliary: auxiliary of time	4.928261
Status of the participle / auxiliary: unclassifiable <i>zijn</i>	1.644758
Status of the participle / auxiliary: passive <i>zijn</i>	3.129873
Status of the participle / auxiliary: <i>worden</i>	5.125976

Variance Inflation Factors

Predictors	V.I.F.
Syntactic persistence: preceding [AUX+PART] order	1.637317
Syntactic persistence: no preceding cluster	1.616749
Frequency of the participle	1.074594
Accent distribution (begin): 0 to 1 syllables	3.307576
Accent distribution (begin): 2 to 3 syllables	3.019115
Accent distribution (end): 0 to 1 syllables	4.572808
Accent distribution (end): 2 to 3 syllables	2.773583
Extraposited constituent: complement of V	1.437830
Extraposited constituent: complement of preverbal N, A	1.206984

Summing up

- We have been able to detect and weight the set of variables that governs the choice of word order:
 - Semantic, discursive, morpho-syntactic features determine choice of word order *simultaneously* (while controlling for language-external variables)
 - The statistical model describes, explains and predicts the variation to a reasonable extent (but, obviously, progress is still possible)

Discussion



The coexistence question

- Why do the variables have the effect they have? How can we **interpret** the results of the multivariate analysis in a more linguistic way?
- Can some of the variables be subsumed under a more **general principle** (e.g. topic-focus)?
- What is the **cognitive-functional difference** between both word order variants? Why do both [AUX+PART] and [PART+AUX] coexist?

The coexistence question

- Most of the principles that have been proposed for interconstituent variation do not seem to apply to intraconstituent variation types:
 - Avoid ambiguity
 - Mark pragmatic value
 - Postpone complex constituents
- Psycholinguistic explanation (processing efficiency)
- However:
 - Embryonic (hypothesis to be tested)

The coexistence question

- E.g., the effect of participial frequency
 - STEP 1: [AUX+PART] word order is dominant when the participle is highly frequent
 - STEP 2: More frequent items are more entrenched and more accessible, and are therefore processed more easily
 - STEP 3: Hence, [AUX+PART] word order is chosen in circumstances where language users experience no difficulty in retrieving and processing the participle
 - STEP 4: Mutatis mutandis, [PART+AUX] word order is chosen in circumstances where language users experience more difficulties in retrieving and processing the participle

Methodological and theoretical issues

- Are there theoretical models that account for this type of multivariate model?

Methodological and theoretical issues

- Are there theoretical models that account for this type of multivariate model?
- Within QLVL
 - Kris Heylen: order of verb arguments in German middle field
 - Jose Tummers: inflection of prenominal adjectives in Dutch
 - Stefan Grondelaers: presentative *er* in Dutch
- Outside QLVL
 - Stefan Gries: particle placement
 - Stefanie Wulff: order of prenominal adjectives
 - Holger Diessel: order of main and adverbial clauses

Methodological and theoretical issues

- Are there theoretical models that account for this type of multivariate model?
- Usage-based approaches to language (e.g. Barlow & Kemmer 1999)
- Linguistic system is grounded in usage events:
 - The linguistic system is formed by usage data
 - Usage data follow from and (re)shape linguistic system

Methodological and theoretical issues

- Are there theoretical models that account for this type of multivariate model?
- What are the features to include in such a model?
 - Towards a more principled selection of exp. variables?
 - Inclusion of language-external variables?



For further information:

gert.desutter@arts.kuleuven.be

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

Prediction: classification table

		Predicted responses	
		[AUX+PART] order	[PART+AUX] order
Observed responses	[AUX+PART] order	1472 (61.59%)	129 (5.40%)
	[PART+AUX] order	411 (17.20%)	378 (15.82%)

- Correctly predicted responses: 77.41% (intercept only = 66.99%)
- Drawbacks (Harrell on *R-help*):
 - Arbitrary cut-off (the proportion classified correctly is arbitrary)
 - Improper scoring rule (not optimized when model estimates true probabilities of events)
 - The proportion classified correctly can decrease when an important variable is added to the model