Variable selection in Logistic Regression:

The British English dative alternation

Daphne Theijssen

PhD student
Department of Linguistics
Radboud University Nijmegen
d.theijssen@let.ru.nl
http://lands.let.ru.nl/∼daphne

ESSLLI'09
Bordeaux
July 20-31, 2009

Introduction

Dative alternation

- *The evil queen gives Snowwhite the poisonous apple.*
  - Double object construction
- *The evil queen gives the poisonous apple to Snowwhite.*
  - Prepositional dative construction

Introduction

Dative alternation

- *The evil queen gives Snowwhite the poisonous apple.*
  - Double object construction
- *The evil queen gives the poisonous apple to Snowwhite.*
  - Prepositional dative construction

Introduction

Dative alternation

- *The evil queen gives Snowwhite the poisonous apple.*
  - Double object construction
- *The evil queen gives the poisonous apple to Snowwhite.*
  - Prepositional dative construction

Introduction

Dative alternation

Dative alternation

Well-studied, for example

- Syntactic approach (e.g. Quirk et al. 1972)
- Semantic approach (e.g. Gries and Stefanowitsch 2004)
- Discourse approach (e.g. Collins 1995)

Introduction

Dative alternation

Dative alternation

Well-studied, for example

- Syntactic approach (e.g. Quirk et al. 1972)
- Semantic approach (e.g. Gries and Stefanowitsch 2004)
- Discourse approach (e.g. Collins 1995)

Introduction

Aim of the project

- Integrate the factors proposed in the literature in one model that
  - can predict which construction is used
  - is interpretable
- Use different techniques:
  - **Logistic Regression**
  - Maximum Entropy
  - Bayesian Networks
- Evaluate suitability for British English dative alternation (and similar phenomena)

Introduction

Aim of the project

- Integrate the factors proposed in the literature in one model that
  - can predict which construction is used
  - is interpretable
- Use different techniques:
  - **Logistic Regression**
  - Maximum Entropy
  - Bayesian Networks
- Evaluate suitability for British English dative alternation (and similar phenomena)

Introduction

Aim of the project

- Integrate the factors proposed in the literature in one model that
  - can predict which construction is used
  - is interpretable

- Use different techniques:
  - **Logistic Regression**
  - Maximum Entropy
  - Bayesian Networks

- Evaluate suitability for British English dative alternation (and similar phenomena)

Introduction

Variable selection in logistic regression

Three steps:

1. **variable selection**
2. establishing the coefficients (weights) for the features
3. evaluating the model

Approaches in variable selection

- all (significant) features (e.g. Bresnan et al. 2007)
- stepwise forward selection (e.g. Grondelaers and Speelman 2007)
- stepwise backward selection (e.g. Blackwell 2005)
- (trying all combinations)

Introduction

Variable selection in logistic regression

Three steps:

1. **variable selection**
2. establishing the coefficients (weights) for the features
3. evaluating the model

Approaches in variable selection

- all (significant) features (e.g. Bresnan et al. 2007)
- stepwise forward selection (e.g. Grondelaers and Speelman 2007)
- stepwise backward selection (e.g. Blackwell 2005)
- (trying all combinations)

Introduction

Variable selection in logistic regression

Which method should we use?

British English data

Data set

- Employ corpus with syntactic annotations: ICE-GB
- Extracted 915 instances

	nr of instances		
Medium	d.obj.	pr.dat.	Total
Spoken	399	151	550
Written	263	102	365
Total	662	253	915

British English data

Data set

- Employ corpus with syntactic annotations: ICE-GB
- Extracted 915 instances

	nr of instances		
Medium	d.obj.	pr.dat.	Total
Spoken	399	151	550
Written	263	102	365
Total	662	253	915

British English data

Features

Manual annotation for 14 features:

recipient

Animacy
Definiteness
Discourse givenness
Number
Person
Pronominality

theme

Concreteness
Definiteness
Discourse givenness
Number
Pronominality

Length difference
Semantic verb class
Structure parallellism

British English data

Features

Manual annotation for 14 features:

 *recipient*

Animacy
Definiteness
Discourse givenness
Number
Person
Pronominality

 *theme*

Concreteness
Definiteness
Discourse givenness
Number
Pronominality

Length difference
Semantic verb class
Structure parallellism

British English data

Features

Add medium to feature list, as well as all interactions with it:

1. Animacy of recipient
2. Concreteness of theme
3. Definiteness of recipient
4. Definiteness of theme
5. Discourse givenness of recipient
6. Discourse givenness of theme
7. Number of recipient
8. Number of theme
9. Person of recipient
10. Pronominality of recipient
11. Pronominality of theme
12. Length difference
13. Semantic verb class
14. Structure parallellism
15. Medium
16. Animacy of recipient:Medium
17. Concreteness of theme:Medium
18. Definiteness of recipient:Medium
19. Definiteness of theme:Medium
20. Discourse givenness of recipient:Medium
21. Discourse givenness of theme:Medium
22. Number of recipient:Medium
23. Number of theme:Medium
24. Person of recipient:Medium
25. Pronominality of recipient:Medium
26. Pronominality of theme:Medium
27. Length difference:Medium
28. Semantic verb class:Medium
29. Structure parallellism:Medium

British English data

Features

Add medium to feature list, as well as all interactions with it:

1. Animacy of recipient
2. Concreteness of theme
3. Definiteness of recipient
4. Definiteness of theme
5. Discourse givenness of recipient
6. Discourse givenness of theme
7. Number of recipient
8. Number of theme
9. Person of recipient
10. Pronominality of recipient
11. Pronominality of theme
12. Length difference
13. Semantic verb class
14. Structure parallellism
15. Medium
16. Animacy of recipient:Medium
17. Concreteness of theme:Medium
18. Definiteness of recipient:Medium
19. Definiteness of theme:Medium
20. Discourse givenness of recipient:Medium
21. Discourse givenness of theme:Medium
22. Number of recipient:Medium
23. Number of theme:Medium
24. Person of recipient:Medium
25. Pronominality of recipient:Medium
26. Pronominality of theme:Medium
27. Length difference:Medium
28. Semantic verb class:Medium
29. Structure parallellism:Medium

British English data

Verb bias

- towards double object (e.g. *give*)
- towards prepositional dative (e.g. *pay*)
- no clear bias (e.g. *lend*)

Include verb + its semantic class (i.e. verb sense) as random effect

British English data

Verb bias

- towards double object (e.g. *give*)
- towards prepositional dative (e.g. *pay*)
- no clear bias (e.g. *lend*)

Include verb + its semantic class (i.e. verb sense) as random effect

British English data

Verb bias

- towards double object (e.g. *give*)
- towards prepositional dative (e.g. *pay*)
- no clear bias (e.g. *lend*)

Include verb + its semantic class (i.e. verb sense) as random effect

Method

Comparing the selection methods

- 29 features, mostly binary
- two types of models:
  - mixed models (with verb sense as random effect)
  - models without a random effect
- 3 different variable selection procedures
  - stepwise forward selection
  - stepwise backward selection
  - including all (significant)
- **Why not build all six** and compare them with respect to
  - their prediction accuracy
  - their interpretability

Method

Comparing the selection methods

- 29 features, mostly binary
- two types of models:
  - mixed models (with verb sense as random effect)
  - models without a random effect
- 3 different variable selection procedures
  - stepwise forward selection
  - stepwise backward selection
  - including all (significant)
- **Why not build all six** and compare them with respect to
  - their prediction accuracy
  - their interpretability

Method

Comparing the selection methods

- 29 features, mostly binary
- two types of models:
  - mixed models (with verb sense as random effect)
  - models without a random effect
- 3 different variable selection procedures
  - stepwise forward selection
  - stepwise backward selection
  - including all (significant)
- **Why not build all six** and compare them with respect to
  - their prediction accuracy
  - their interpretability

Method

Comparing the selection methods

- 29 features, mostly binary
- two types of models:
  - mixed models (with verb sense as random effect)
  - models without a random effect
- 3 different variable selection procedures
  - stepwise forward selection
  - stepwise backward selection
  - including all (significant)
- **Why not build all six** and compare them with respect to
  - their prediction accuracy
  - their interpretability

Method

Some details

$$\log odds(C_{ij} = 1) = \alpha + \sum_{k=1}^{n}(\beta_k V_{ijk}) \quad (+r_j)$$

- coefficients α, β_k and r_j found with Maximum Likelihood Estimation (MLE)
- Evaluation measures:
  - the number of features selected
  - prediction accuracy when train=test
  - prediction accuracy in 10-fold cross-validation
  - the coefficients (interpretability)

Method

Some details

$$\log odds(C_{ij} = 1) = \alpha + \sum_{k=1}^{n}(\beta_k V_{ijk}) \quad (+r_j)$$

- coefficients α, β_k and r_j found with Maximum Likelihood Estimation (MLE)

- Evaluation measures:
  - the number of features selected
  - prediction accuracy when train=test
  - prediction accuracy in 10-fold cross-validation
  - the coefficients (interpretability)

Method

Some details

$$\log odds(C_{ij} = 1) = \alpha + \sum_{k=1}^{n}(\beta_k V_{ijk}) \quad (+r_j)$$

- coefficients α, β_k and r_j found with Maximum Likelihood Estimation (MLE)
- Evaluation measures:
  - the number of features selected
  - prediction accuracy when train=test
  - prediction accuracy in 10-fold cross-validation
  - the coefficients (interpretability)

Results

Mixed models

			model fit	*10-fold cv*
selection	#features	baseline	accuracy	aver. acc.
1. significant	5	0.723	0.936	0.825
2. forward	4	0.723	0.938	0.832
3. backward	4	0.723	0.938	0.832

Results

Mixed models

Effect	1. significant		2. forward		3. backward		
th=pronominal; wr	-2.01	*					
length difference	-2.52	***	-2.41	***	-2.41	***	d.obj.
rec=local	-2.68	***	-1.86	***	-1.86	***	↑
rec=given			-1.48	***	-1.48	***	|
th=definite	1.67	***					|
th=pronominal	2.15	***					↓
th=given			2.32	***	2.32	***	p.dat.
(intercept)	1.27	**	2.53	***	2.53	***	

Results

Models without a random effect

selection	#features	baseline	*model fit* accuracy	*10-fold cv* aver. acc.
1. significant	5	0.723	0.882	0.882
2. forward	5	0.723	0.883	0.870
3. backward	8	0.723	0.882	0.875

Results

Models without a random effect

Effect	1. significant		2. forward		3. backward		
length difference; sp					-2.29	***	
rec=pronominal; wr					-2.07	***	d.obj.
length difference	-1.75	***	-1.97	***			↑
rec=pronominal; sp					-1.71	***	
length difference; wr					-1.49	***	
rec=definite			-1.18	***	-1.20	***	
rec=local	-1.13	***	-1.20	***			
rec=pronominal	-1.15	***	-1.20	***			
th=definite			1.12	***			
(intercept)					1.37	***	↓
th=given			0.95	**	1.44	***	p.dat.
th=concrete	1.50	***	1.52	***	1.27	***	

Conclusion

- With our medium-sized data set, six approaches have led to five different models
- All agree with effects found by other researchers: pronominal, relatively short, local, discourse given, definite and concrete objects typically precede objects with the opposite characteristics
- The mixed models yield a better model fit but they do not generalize as well as the models without a random effect
- Data set too small?

Conclusion

- With our medium-sized data set, six approaches have led to five different models
- All agree with effects found by other researchers: pronominal, relatively short, local, discourse given, definite and concrete objects typically precede objects with the opposite characteristics

- The mixed models yield a better model fit but they do not generalize as well as the models without a random effect
- Data set too small?

Conclusion

- Which method is best depends on one's goal
- Researchers should be more explicit about their choices

- Any suggestions are welcome

Conclusion

- Which method is best depends on one's goal
- Researchers should be more explicit about their choices

- Any suggestions are welcome

Thank you!

Daphne Theijssen
PhD student
Department of Linguistics
Radboud University Nijmegen
d.theijssen@let.ru.nl
http://lands.let.ru.nl/∼daphne