Active Learning and the Irish Treebank

Teresa Lynn, Jennifer Foster, Mark Dras,
Elaine Uí Dhonnchadha

Centre for Language Technology, Macquarie University
National Centre for Language Technology, Dublin City University
Centre for Language and Communication Studies, Trinity College Dublin

2nd July 2013

Outline

Irish Language

Treebanking

Active Learning Experiments

Conclusion

Outline

Irish Language

Treebanking

Active Learning Experiments

Conclusion

Irish — a low density language

- Official EU language
  - VSO Celtic Language
  - inflected prepositions: *agam* 'at me'; *agat* 'at you'
  - progressive aspectual phrases: *tá sé ag rith* 'he is running'
  - clefting: *is mise atá ag caint* 'it's me who is talking'
- National Corpus of Ireland - Irish (NCII):
  - 30 million word corpus - Foras na Gaeilge
- Morphological Analyser; POS-tagger; Shallow chunking parser
  - Elaine Uí Dhonnchadha (Trinity College Dublin)
- Parallel Data
  - Kevin Scannell (St. Louis University, Missouri)
  - Traslan Teoranta (data made available to NCLT, Dublin)
- META-NET. White paper published in 2012:
  "The Irish Language in the Digital Age"
  ...more resources needed

Irish — a low density language

- Official EU language
  - VSO Celtic Language
  - inflected prepositions: *agam* ‘at me’; *agat* ‘at you’
  - progressive aspectual phrases: *tá sé ag rith* ‘he is running’
  - clefting: *is mise atá ag caint* ‘it’s me who is talking’
- National Corpus of Ireland - Irish (NCII):
  - 30 million word corpus - Foras na Gaeilge
- Morphological Analyser; POS-tagger; Shallow chunking parser
  - Elaine Uí Dhonnchadha (Trinity College Dublin)
- Parallel Data
  - Kevin Scannell (St. Louis University, Missouri)
  - Traslan Teoranta (data made available to NCLT, Dublin)
- META-NET. White paper published in 2012:
  ”The Irish Language in the Digital Age”
  . . . more resources needed

Irish — a low density language

- Official EU language
  - VSO Celtic Language
  - inflected prepositions: *agam* 'at me'; *agat* 'at you'
  - progressive aspectual phrases: *tá sé ag rith* 'he is running'
  - clefting: *is mise atá ag caint* 'it's me who is talking'
- National Corpus of Ireland - Irish (NCII):
  - 30 million word corpus - Foras na Gaeilge
- Morphological Analyser; POS-tagger; Shallow chunking parser
  - Elaine Uí Dhonnchadha (Trinity College Dublin)
- Parallel Data
  - Kevin Scannell (St. Louis University, Missouri)
  - Traslan Teoranta (data made available to NCLT, Dublin)
- META-NET. White paper published in 2012:
  "The Irish Language in the Digital Age"
  ... more resources needed

Irish — a low density language

- Official EU language
  - VSO Celtic Language
  - inflected prepositions: *agam* 'at me'; *agat* 'at you'
  - progressive aspectual phrases: *tá sé ag rith* 'he is running'
  - clefting: *is mise atá ag caint* 'it's me who is talking'
- National Corpus of Ireland - Irish (NCII):
  - 30 million word corpus - Foras na Gaeilge
- Morphological Analyser; POS-tagger; Shallow chunking parser
  - Elaine Uí Dhonnchadha (Trinity College Dublin)
- Parallel Data
  - Kevin Scannell (St. Louis University, Missouri)
  - Traslan Teoranta (data made available to NCLT, Dublin)
- META-NET. White paper published in 2012: "The Irish Language in the Digital Age" . . . more resources needed

Irish — a low density language

- Official EU language
  - VSO Celtic Language
  - inflected prepositions: *agam* 'at me'; *agat* 'at you'
  - progressive aspectual phrases: *tá sé ag rith* 'he is running'
  - clefting: *is mise atá ag caint* 'it's me who is talking'
- National Corpus of Ireland - Irish (NCII):
  - 30 million word corpus - Foras na Gaeilge
- Morphological Analyser; POS-tagger; Shallow chunking parser
  - Elaine Uí Dhonnchadha (Trinity College Dublin)
- Parallel Data
  - Kevin Scannell (St. Louis University, Missouri)
  - Traslan Teoranta (data made available to NCLT, Dublin)
- META-NET. White paper published in 2012:
  "The Irish Language in the Digital Age"
  . . . more resources needed

Outline

Irish Language

Treebanking

Active Learning Experiments

Conclusion

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set $<$ 1500 sentences $=$ reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set < 1500 sentences $=$ reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set $<$ 1500 sentences $=$ reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set < 1500 sentences = reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set $<$ 1500 sentences $=$ reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Treebanks

General Issues

- Training data for statistical parsers
- Phrase Structure vs Dependencies
- What is a sufficient size?
- Parsing experiments on 13 treebanks (Nivre, 2008):
  - training set $<$ 1500 sentences $=$ reasonably accurate parsing models (Arabic, Slovene)

Key Points of Talk

- Update on development of Irish Treebank
- Extent to which Active Learning can play a role
  - Bootstrapping (Passive Learning, Active Learning)

Irish Dependency Treebank (Lynn et al., 2012)

Built upon existing NLP resources:

- Morphological analyser, POS-tagger, 225 chunked sentences (Uí Dhonnchadha, 2009, PhD Thesis)
- LFG-inspired Dependency Scheme (Çetinoğlu et al., 2010)
- Small amount of Irish LFG research (e.g. Sulger, 2009)
- Head-modifier dependency relations
  - CoNLL-X format — Form; Lemma; CPOS; FPOS; Head; Label

Irish Dependency Treebank

Preliminary parsing experiments

- MaltParser (Nivre et al., 2006)
- Seed set of 300 gold manually annotated sentences
- 10-fold cross validation
- Test variety of feature models
- Select baseline parsing model

Model	**LAS**	**UAS**
Form+POS:	60.6	70.3
Lemma+POS:	61.3	70.8
Form+Lemma+POS:	61.5	70.8
Form+CPOS:	62.1	72.5
Form+Lemma+CPOS:	62.9	72.6
Form+CPOS+POS:	63.0	72.9
Lemma+CPOS+POS:	63.1	72.4
Lemma+CPOS:	63.3	72.7
Form+Lemma+CPOS+POS:	63.3	73.1

IAA experiments

Why the need for Inter-annotator agreement measure?

- Are annotators' judgements consistent/ reliable/ trustworthy?
- Indicates usefulness of data
- Indicates replicability (e.g. clinical studies – diagnoses)
- Analysis of disagreements can be used to improve labelling scheme
- Identifies gaps in the annotation guide

How do we measure IAA?

- Kappa coefficient of agreement (Cohen, 1960)
- Labelled Attachment Score (LAS)
- Unlabelled Attachment Score (UAS)

IAA experiments

Why the need for Inter-annotator agreement measure?

- Are annotators' judgements consistent/ reliable/ trustworthy?
- Indicates usefulness of data
- Indicates replicability (e.g. clinical studies – diagnoses)
- Analysis of disagreements can be used to improve labelling scheme
- Identifies gaps in the annotation guide

How do we measure IAA?

- Kappa coefficient of agreement (Cohen, 1960)
- Labelled Attachment Score (LAS)
- Unlabelled Attachment Score (UAS)

IAA experiments

Why the need for Inter-annotator agreement measure?

- Are annotators' judgements consistent/ reliable/ trustworthy?
- Indicates usefulness of data
- Indicates replicability (e.g. clinical studies – diagnoses)
- Analysis of disagreements can be used to improve labelling scheme
- Identifies gaps in the annotation guide

How do we measure IAA?

- Kappa coefficient of agreement (Cohen, 1960)
- Labelled Attachment Score (LAS)
- Unlabelled Attachment Score (UAS)

IAA experiments

Why the need for Inter-annotator agreement measure?

- Are annotators' judgements consistent/ reliable/ trustworthy?
- Indicates usefulness of data
- Indicates replicability (e.g. clinical studies – diagnoses)
- Analysis of disagreements can be used to improve labelling scheme
- Identifies gaps in the annotation guide

How do we measure IAA?

- Kappa coefficient of agreement (Cohen, 1960)
- Labelled Attachment Score (LAS)
- Unlabelled Attachment Score (UAS)

IAA experiments

IAA-1 (Lynn et al., 2012)

- Two annotators
  - Irish speaking/ linguistic background
- Practice run on 30 random sentences
  - introduce 2nd annotator to labelling scheme and guide
- 50 random sentences - without consultation
- Results:

Kappa (labels)	LAS	UAS
0.7902	74.37%	85.16%

IAA experiments

IAA-1 (Lynn et al., 2012)

- Two annotators
  - Irish speaking/ linguistic background
- Practice run on 30 random sentences
  - introduce 2nd annotator to labelling scheme and guide
- 50 random sentences - without consultation
- Results:

Kappa (labels)	LAS	UAS
0.7902	74.37%	85.16%

IAA experiments

IAA-1 (Lynn et al., 2012)

- Two annotators
  - Irish speaking/ linguistic background
- Practice run on 30 random sentences
  - introduce 2nd annotator to labelling scheme and guide
- 50 random sentences - without consultation
- Results:

Kappa (labels)	LAS	UAS
0.7902	74.37%	85.16%

IAA experiments

IAA-1 (Lynn et al., 2012)

- Two annotators
  - Irish speaking/ linguistic background
- Practice run on 30 random sentences
  - introduce 2nd annotator to labelling scheme and guide
- 50 random sentences - without consultation
- Results:

Kappa (labels)	**LAS**	**UAS**
0.7902	74.37%	85.16%

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments - present study

Post IAA-1 Workshop Analysis

- Both annotators and Irish syntactician
- Compared both annotators' files
- Analysed types of disagreements
- Updated labelling scheme
- Updated annotation guide
- Updated the treebank

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	LAS	UAS
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	LAS	UAS
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	LAS	UAS
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	LAS	UAS
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	LAS	UAS
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	**LAS**	**UAS**
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA experiments

IAA-2

- Same two annotators
- Newly updated labelling scheme and annotation guide
- Practice run on 20 random sentences
- Measured agreement on different set of 50 random sentences:

	Kappa (labels)	**LAS**	**UAS**
IAA-1	0.7902	74.37%	85.16%
IAA-2	0.8463	79.17%	87.75%

- Improved results show benefits of post IAA-1 workshops

IAA-1 analysis

Overview of Types of Disagreements

- interpretation disagreements
  (e.g. legislative text > 200 tokens, obscure terminology)
- human error
- gaps in annotation guide
  (e.g. lack of examples, not clearly described etc.)
- outstanding issues
  (e.g unresolved or newly encountered linguistic phenomena)

IAA-1 analysis

Overview of Types of Disagreements

- interpretation disagreements
  (e.g. legislative text > 200 tokens, obscure terminology)
- human error
- gaps in annotation guide
  (e.g. lack of examples, not clearly described etc.)
- outstanding issues
  (e.g unresolved or newly encountered linguistic phenonema)

IAA-1 analysis

Overview of Types of Disagreements

- interpretation disagreements
  (e.g. legislative text $>$ 200 tokens, obscure terminology)
- human error
- gaps in annotation guide
  (e.g. lack of examples, not clearly described etc.)
- outstanding issues
  (e.g unresolved or newly encountered linguistic phenonema)

IAA-1 analysis

Overview of Types of Disagreements

- interpretation disagreements
  (e.g. legislative text $>$ 200 tokens, obscure terminology)
- human error
- gaps in annotation guide
  (e.g. lack of examples, not clearly described etc.)
- outstanding issues
  (e.g unresolved or newly encountered linguistic phenonema)

IAA-1 analysis

Overview of Types of Disagreements

- interpretation disagreements
  (e.g. legislative text $>$ 200 tokens, obscure terminology)
- human error
- gaps in annotation guide
  (e.g. lack of examples, not clearly described etc.)
- outstanding issues
  (e.g unresolved or newly encountered linguistic phenonema)

IAA-1 analysis: Changes in the labelling scheme

- `xcomp` $->$ `pred` hierarchy:
  - `npred`: nominal predicate
  - `ppred`: prepositional predicate
  - `advpred`: adverbial predicate
  - `adjpred`: adjectival predicate
- `relparticle` $->$ `cleftparticle`:

	npred	cleftparticle	subj	subj	advadjunct
Is	ise	a	chonaic	mé	inné
COP	*she*	*REL*	*saw*	*I*	*yesterday*

‘It is she whom I saw yesterday’

IAA-1 analysis: Changes in the labelling scheme

- `xcomp` −> `pred` hierarchy:
  - `npred`: nominal predicate
  - `ppred`: prepositional predicate
  - `advpred`: adverbial predicate
  - `adjpred`: adjectival predicate
- `relparticle` −> `cleftparticle`:

	npred	cleftparticle	subj	subj	advadjunct
Is	ise	a	chonaic	mé	inné
COP	*she*	*REL*	*saw*	*I*	*yesterday*

‘It is she whom I saw yesterday’

IAA-1 analysis

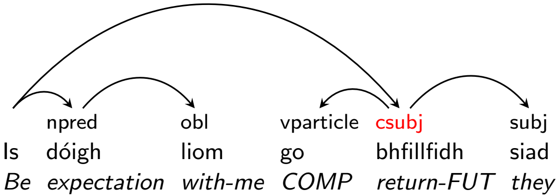
- `adjunct` $->$ `obl`
- `subj` $->$ `csubj`

	npred	obl	vparticle	csubj	subj
Is	dóigh	liom	go	bhfillfidh	siad
Be	*expectation*	*with-me*	*COMP*	*return-FUT*	*they*

'I expect they will return'

IAA-1 analysis

- `adjunct` $->$ `obl`
- `subj` $->$ `csubj`



'I expect they will return'

Re-running parsing experiments

- LAS-1/ UAS-1: original treebank
- LAS-2/ UAS-2: post-workshop updated treebank

Model	LAS-1	UAS-1	LAS-2	UAS-2
Form+POS:	60.6	70.3	64.4	74.2
Lemma+POS:	61.3	70.8	64.6	74.3
Form+Lemma+POS:	61.5	70.8	64.6	74.5
Form+CPOS:	62.1	72.5	65.0	76.1
Form+Lemma+CPOS:	62.9	72.6	66.1	76.2
Form+CPOS+POS:	63.0	72.9	66.0	76.0
Lemma+CPOS+POS:	63.1	72.4	66.0	76.2
Lemma+CPOS:	63.3	72.7	65.1	75.7
Form+Lemma+CPOS+POS:	63.3	73.1	66.5	76.3

Outline

Irish Language

Treebanking

Active Learning Experiments

Conclusion

Overview of Active Learning

General

- Speed up treebank creation — bootstrapping approach
- Passive vs Active
  - Passive — select next set of parse trees for manual correction
  - Active — select 'problematic' parses for manual correction
- Active Learning previously used in NLP:
  - information extraction (Scheffer et al., 2001)
  - text categorisation (Lewis and Gale, 1994)
  - word-sense disambiguation (Chen et al., 2006)
  - parsing (Osborne and Baldridge, 2004)

Useful for building a resource?

- Some evidence of this with Interlinear Glossed Texts (Baldridge and Palmer, 2009)
- Treebanks?

Overview of Active Learning

General

- Speed up treebank creation — bootstrapping approach
- Passive vs Active
  - Passive — select next set of parse trees for manual correction
  - Active — select 'problematic' parses for manual correction
- Active Learning previously used in NLP:
  - information extraction (Scheffer et al., 2001)
  - text categorisation (Lewis and Gale, 1994)
  - word-sense disambiguation (Chen et al., 2006)
  - parsing (Osborne and Baldridge, 2004)

Useful for building a resource?

- Some evidence of this with Interlinear Glossed Texts (Baldridge and Palmer, 2009)
- Treebanks?

Overview of Active Learning

General

- Speed up treebank creation — bootstrapping approach
- Passive vs Active
  - Passive — select next set of parse trees for manual correction
  - Active — select 'problematic' parses for manual correction
- Active Learning previously used in NLP:
  - information extraction (Scheffer et al., 2001)
  - text categorisation (Lewis and Gale, 1994)
  - word-sense disambiguation (Chen et al., 2006)
  - parsing (Osborne and Baldridge, 2004)

Useful for building a resource?

- Some evidence of this with Interlinear Glossed Texts (Baldridge and Palmer, 2009)
- Treebanks?

Overview of Active Learning

General

- Speed up treebank creation — bootstrapping approach
- Passive vs Active
  - Passive — select next set of parse trees for manual correction
  - Active — select 'problematic' parses for manual correction
- Active Learning previously used in NLP:
  - information extraction (Scheffer et al., 2001)
  - text categorisation (Lewis and Gale, 1994)
  - word-sense disambiguation (Chen et al., 2006)
  - parsing (Osborne and Baldridge, 2004)

Useful for building a resource?

- Some evidence of this with Interlinear Glossed Texts (Baldridge and Palmer, 2009)
- Treebanks?

How to identify 'problematic' (informative) parses?

Query By Uncertainty (QBU) - (Cohn et al., 1994)

- select for manual correction when learner is least confident

Query By Committee (QBC) - (Seung et al., 1992)

- select based on disagreement among committee of learners

Bootstrapping Algorithm

$t \leftarrow$ seed training set
Train a parsing model, p, using the trees in t
repeat
  $u \leftarrow$ a set of X unlabelled sentences
  Parse u with p to yield u_p
  $u\prime \leftarrow$a subset of Y sentences from u
  Hand-correct $u\prime_p$ to yield $u\prime_{gold}$
  $t \leftarrow t + u\prime_{gold}$ {Add $u\prime_{gold}$ to t}
  Train a parsing model, p, using the trees in t
until convergence

$u = 200$
$u\prime = 50$

Bootstrapping Algorithm (Passive vs Active)

The difference in versions lies in how $u\prime$ is chosen.

Passive

$u\prime$ is random selection of 50 sentences from u.

Active - (QBC)

$u\prime$ is top 50 most 'disagreed upon' sentences from u.

Active Learning Experiment Setup

Details

- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee = MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Active Learning Experiment Setup

Details

- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee = MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Active Learning Experiment Setup

Details

- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee = MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Active Learning Experiment Setup

Details

- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee = MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Active Learning Experiment Setup

Details

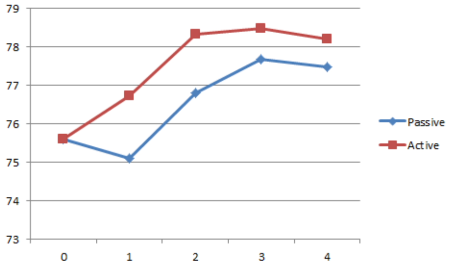
- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee =
  MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Active Learning Experiment Setup

Details

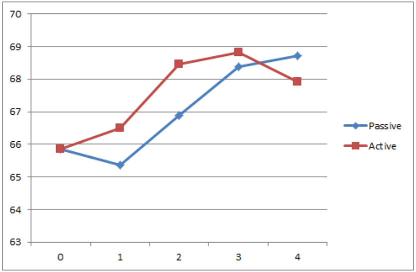
- 450 Gold Trees = seed training (150); development set (150); test set (150)
- Two versions of bootstrapping algorithm: Active & Passive
- Four iterations (50 sentences manually corrected each time)
- Active Committee =
  MaltParser (Nivre et al., 2006) & Mate parser (Bohnet, 2010)
- Disagreement between two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

Results: Passive vs Active (UAS)



x-axis= the number of training iterations
y-axis = unlabelled attachment score

Results: Passive vs Active (LAS)



x-axis= the number of training iterations
y-axis = labelled attachment score

Differences between Active and Passive sentences

	It. 1	It.2	It.3	It.4
	Average Sentence Length			
Passive	18.6	28.6	23.9	24.5
Active	18.8	25.5	24.8	35.9
	Correction Effort			
Passive	23.8	30.2	27.0	23.8
Active	36.7	37.6	32.4	32.8

- Correction => gold parse trees
- Correction effort = disagreement between the automatic parse and its correction (1-LAS)

- Modest gains - worth the effort?

Differences between Active and Passive sentences

	It. 1	It.2	It.3	It.4
	Average Sentence Length			
Passive	18.6	28.6	23.9	24.5
Active	18.8	25.5	24.8	35.9
	Correction Effort			
Passive	23.8	30.2	27.0	23.8
Active	36.7	37.6	32.4	32.8

- Correction => gold parse trees
- Correction effort = disagreement between the automatic parse and its correction (1-LAS)
- Modest gains - worth the effort?

Outline

Irish Language

Treebanking

Active Learning Experiments

Conclusion

Summary

- Used IAA to finalise annotation scheme
- Bootstrapping parser => bootstrapping the treebank
- Explored the role of AL in treebank development
  - How will this picture change over time?
- Improved parsing results (based on test set):
  Baseline = **LAS 63.4%**
  Final Active Model = **LAS 68.0%**
  Final Passive Model = **LAS 67.2%**

Summary

- Used IAA to finalise annotation scheme
- Bootstrapping parser => bootstrapping the treebank
- Explored the role of AL in treebank development
  - How will this picture change over time?
- Improved parsing results (based on test set):
  Baseline = **LAS 63.4%**
  Final Active Model = **LAS 68.0%**
  Final Passive Model = **LAS 67.2%**

Summary

- Used IAA to finalise annotation scheme
- Bootstrapping parser => bootstrapping the treebank
- Explored the role of AL in treebank development
  - How will this picture change over time?
- Improved parsing results (based on test set):
  Baseline = **LAS 63.4%**
  Final Active Model = **LAS 68.0%**
  Final Passive Model = **LAS 67.2%**

Summary

- Used IAA to finalise annotation scheme
- Bootstrapping parser => bootstrapping the treebank
- Explored the role of AL in treebank development
  - How will this picture change over time?
- Improved parsing results (based on test set):
  Baseline = **LAS 63.4%**
  Final Active Model = **LAS 68.0%**
  Final Passive Model = **LAS 67.2%**

Future Work

- Try new Active Learning configurations
  - e.g. Swapping parsers - use Mate parser as main parser
- What sentence length is too long?
- More annotation
- Parsing experiments
  - Self-training and Co-training
  - Unsupervised parsing

Future Work

- Try new Active Learning configurations
  - e.g. Swapping parsers - use Mate parser as main parser
- What sentence length is too long?
- More annotation
- Parsing experiments
  - Self-training and Co-training
  - Unsupervised parsing

Future Work

- Try new Active Learning configurations
  - e.g. Swapping parsers - use Mate parser as main parser
- What sentence length is too long?
- More annotation
- Parsing experiments
  - Self-training and Co-training
  - Unsupervised parsing

Future Work

- Try new Active Learning configurations
  - e.g. Swapping parsers - use Mate parser as main parser
- What sentence length is too long?
- More annotation
- Parsing experiments
  - Self-training and Co-training
  - Unsupervised parsing

Future Work

- Try new Active Learning configurations
  - e.g. Swapping parsers - use Mate parser as main parser
- What sentence length is too long?
- More annotation
- Parsing experiments
  - Self-training and Co-training
  - Unsupervised parsing

Go raibh maith agaibh
Thank you
Merci