Active Learning and the Irish Treebank

Teresa Lynn[1,2], **Jennifer Foster**[1], **Mark Dras**[2] and **Elaine Uí Dhonnchadha**[3]
[1]NCLT/CNGL, Dublin City University, Ireland
[2]Department of Computing, Macquarie University, Sydney, Australia
[3]Centre for Language and Communication Studies, Trinity College, Dublin, Ireland
[1]{tlynn,jfoster}@computing.dcu.ie
[2]{teresa.lynn,mark.dras}@mq.edu.au, [3]uidhonne@tcd.ie

Abstract

We report on our ongoing work in developing the Irish Dependency Treebank, describe the results of two Inter-annotator Agreement (IAA) studies, demonstrate improvements in annotation consistency which have a knock-on effect on parsing accuracy, and present the final set of dependency labels. We then go on to investigate the extent to which active learning can play a role in treebank and parser development by comparing an active learning bootstrapping approach to a passive approach in which sentences are chosen at random for manual revision. We show that active learning outperforms passive learning, but when annotation effort is taken into account, it is not clear how much of an advantage the active learning approach has. Finally, we present results which suggest that adding automatic parses to the training data along with manually revised parses in an active learning setup does not greatly affect parsing accuracy.

1 Introduction

The Irish language is an official language of the European Union and is the first national language of the Republic of Ireland. It is a Verb-Subject-Object language, belonging to the Celtic language group. Irish is considered a low-density language, lacking in sufficient resources for various natural language processing (NLP) applications. The development of a dependency treebank is part of a recent initiative to address this lack of resources, as has been the case for, for example, Danish (Kromann, 2003), Slovene (Džeroski et al., 2006) and Finnish (Haverinen et al., 2010). Statistical parsers are data-driven and require a sufficient number of parsed sentences to learn from. One of the expected uses of a treebank for Irish is to provide training data for the first Irish statistical dependency parser which will form the basis of useful NLP applications such as Machine Translation or Computer Aided Language Learning.

What counts as a sufficient number of trees for training an Irish statistical dependency parser remains an open question. However, what is clear is that the parser needs to have encountered a linguistic phenomenon in training in order to learn how to accurately analyse it. Creating a treebank is a resource-intensive process which requires extensive linguistic research in order to design an appropriate labelling scheme, as well as considerable manual annotation (parsing). In general, manual annotation is desired to ensure high quality treebank data. Yet, as is often encountered when working with language, the task of manually annotating text can become repetitive, involving frequent encounters with similar linguistic structures.

In an effort to speed up the creation of treebanks, there has been an increased focus towards automating, or at least, semi-automating the process using various bootstrapping techniques. A basic bootstrapping approach such as that outlined by Judge et al. (2006) involves several steps. Firstly a parser is trained on a set of gold standard trees. This parser is then used to parse a new set of unseen sentences. When these new trees are reviewed and corrected, they are combined with the first set of trees and used to train a new parsing model. These steps are repeated until all sentences are parsed. By adding to the training data on each iteration, the parser is expected to improve progressively. The process of correcting

Teresa Lynn, Jennifer Foster, Mark Dras and Elaine Uí Dhonnchadha. 2012. Active Learning and the Irish Treebank. In *Proceedings of Australasian Language Technology Association Workshop*, pages 23−32.

the trees should become, in turn, less onerous. An *active learning* bootstrapping approach, also referred to as selective sampling, focuses on selecting 'informative' sentences on which to train the parser on each iteration. Sentences are regarded as informative if their inclusion in the training data is expected to fill gaps in the parser's knowledge.

This paper is divided into two parts. In Part One, we report on our ongoing work in developing the Irish Dependency Treebank, we describe the results of two Inter-annotator Agreement (IAA) studies and we present the finalised annotation scheme. In Part Two, we assess the extent to which active learning can play a role in treebank and parser development. We compare an active learning bootstrapping approach to a passive one in which sentences are chosen at random for manual revision. We show that we can reach a certain level of parsing accuracy with a smaller training set using active learning but the advantage over passive learning is relatively modest and may not be enough to warrant the extra annotation effort involved.

2 The Irish Dependency Treebank

The work discussed in this paper builds upon previous work on the Irish Dependency Treebank by Lynn et al. (2012). The treebank consists of randomly selected sentences from the National Corpus for Ireland (NCII) (Kilgarriff et al., 2006). This 30 million word corpus comprises text from news sources, books, government legislative acts, websites and other media. A 3,000 sentence gold-standard part-of-speech (POS) tagged corpus was produced by Uí Dhonnchadha et al. (2003). Another 225 hand-crafted Irish sentences are also available as a result of work by Uí Dhonnchadha (2009). These 3,225 sentences, subsequently randomised, formed the starting point for the treebank.

2.1 Inter-annotator agreement experiments

Inter-annotator agreement (IAA) experiments are used to assess the consistency of annotation within a treebank when more than one annotator is involved. As discussed by Artstein and Poesio (2008), an IAA result not only reveals information about the annotators, i.e. consistency and reliability, but it can also identify shortcomings in the annotation scheme or gaps in the annotation guide. The analysis of IAA results can also provide insight as to the types of disagreements involved and their sources.

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

Table 1: IAA results. LAS or Labelled Attachment Score is the percentage of words for which the two annotators have assigned the same head and label. UAS or Unlabelled Attachment Score is the percentage of words for which the two annotators have assigned the same head.

In previous work (Lynn et al., 2012) , an inter-annotator agreement assessment was conducted by selecting 50 sentences at random from the Irish POS-tagged corpus. Two nominated annotators (Irish-speaking linguists) annotated the sentences individually, according to the protocol set out in the annotation guide, without consultation. The results are shown in the first row of Table 1. For this present study, we held three workshops with the same two annotators and one other fluent Irish speaker/linguist to analyse the results of IAA-1. We took both annotators' files from IAA-1 to assess the types of disagreements that were involved. The analysis highlighted many gaps in the annotation guide along with the requirement for additional labels or new analyses. Thus, we updated the scheme and the annotation guide to address these issues. We then carried out a second IAA assessment (IAA-2) on a set of 50 randomly selected sentences. The results are shown in the second row of Table 1. A notable improvement in IAA-2 results demonstrates that the post-IAA-1 analysis, the resulting workshop discussions and the subsequent updates to the annotation scheme and guidelines were highly beneficial steps towards improving the quality of the treebank.

We have reviewed and updated the already annotated trees (300 sentences) to ensure consistency throughout the treebank. In total, 450 gold standard trees are now available. 150 of these sentences have been doubly annotated: prior to IAA-1, we used a set of 30 sentences for discussion/ training purposes to ensure the annotation guide was comprehensible to both annotators. A set of 20 sentences were used for the same purposes prior to IAA-2.

2.2 Sources of annotator disagreements

The analysis of IAA results provided information valuable for the improvement of the annotation scheme. This analysis involved the comparison of both annotators' files of 50 sentences to see where they disagreed and the types of disagreements involved. Close examination of the disagreements allowed us to categorise them as: (i) Interpretation disagreements (ii) Errors (iii) Gaps in annotation guide (iv) Outstanding issues with the dependency scheme.

2.2.1 Interpretation disagreements

The treebank data was extracted from the NCII which contains many examples of Irish legislative text. Some of these sentences are over 200 tokens in length and use obscure terminology or syntactic structures. Both annotators encountered difficulties in (i) interpreting the intended meaning of these sentences and (ii) analysing their structures. Sources of disagreement included long distance dependencies and coordinated structures.

2.2.2 Errors

Human error played a relatively small role as both annotators carried out careful reviews of their annotations. Nevertheless, some discrepancies were due to an annotator applying the wrong label even though they were aware of the correct one.

2.2.3 Gaps in the annotation guide

Gaps relate to a lack of sufficient examples in the annotation guide or lack of coverage for certain structures. For example, our analysis of IAA-1 confusions revealed that differences between the labels `padjunct` (prepositional modifier) and `obl` (oblique) were not described clearly enough.

2.2.4 Outstanding issues in the dependency scheme

We also noted during the workshops that there were still some issues we had yet to resolve. For example, in the earlier labelling scheme, we used the Sulger (2009) analysis to label as `adjunct` the relationship between predicates and prepositional phrases in a copula construction. An example is *Is maith liom tae* 'I like tea' (lit. 'tea is good with me'). However, in such a construction, the prepositional phrase – *liom* 'with me' in this case – is not optional. We choose instead to label them as `obl`. Other outstanding issues involved linguistic phenomena that had not arisen during earlier annotations and thus required discussion at this stage.

The annotation scheme defined by Lynn et al. (2012) is inspired by Lexical Functional Grammar (Bresnan, 2001) and similar to that of Çetinoğlu et al. (2010). As a result of IAA-2, we have extended the scheme by adding a hierarchical structure where appropriate and updating some analyses. The final scheme is presented in Table 2. In what follows we briefly discuss some updates to the scheme.

Labelling of predicates Our prior labelling scheme (Lynn et al., 2012) regarded predicates of both the copula *is* and the substantive verb *bí* as `xcomp` - as inspired by discussions in the LFG literature e.g. Dalrymple et al. (2004), Sulger (2009). However, open complement verbs (infinitive verbs and progressive verb phrases) were also labelled as `xcomp`. In order to differentiate these different kinds of functions, we have adopted a `pred` hierarchy of `npred`, `ppred`, `adjpred` and `advpred`. While a more fine-grained labelling scheme could result in more data sparsity, it also results in a more precise description of Irish syntax. Examples are provided in Figure 1 and Figure 2.

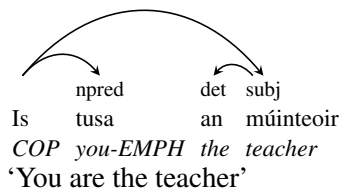

Figure 1: Dependency structure with new nominal predicate labelling (identity copular construction)

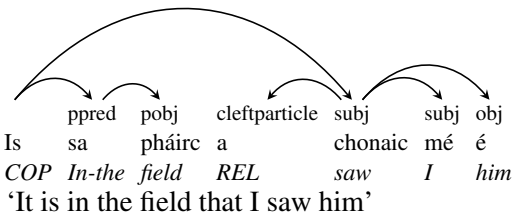

Figure 2: Dependency structure with new prepositional predicate labelling (cleft copular construction)

Cleft constructions - cleft particle Clefting or fronting is a commonly used structure in the Irish

dependency label	function
top	root
punctuation	internal and final punctuation
subj	subject
csubj	clausal subject
obj	object
pobj	object of preposition
vnobj	object of verbal noun
obl	oblique object
obl2	second oblique object
obl_ag	oblique agent
det	determiner
det2	post or pre-determiner
dem	demonstrative pronoun
poss	possessive pronoun
aug	augment pronoun
quant	quantifier
coord	coordinate
relmod	relative modifier
particle	particle
relparticle	relative particle
cleftparticle	cleft particle
advparticle	adverbial particle
nparticle	noun particle
vparticle	verb particle
particlehead	particle head
qparticle	quantifier particle
vocparticle	vocative particle
addr	addressee
adjunct	adjunct
adjadjunct	adjectival modifier
advadjunct	adverbial modifier
nadjunct	nominal modifier
padjunct	prepositional modifier
subadjunct	subordinate conjunction
toinfinitive	infinitive verb marker
app	noun in apposition
xcomp	open complement
comp	closed complement
pred	predicate
ppred	prepositional predicate
npred	nominal predicate
adjpred	adjectival predicate
advpred	adverbial predicate
subj_q	subject (question)
obj_q	object (question)
advadjunct_q	adverbial adjunct (question)
for	foreign (non-Irish) word

Table 2: The Irish Dependency Treebank labels: sub-labels are indicated in bold and their parents in italics

language. Elements are fronted to predicate position to create emphasis. Irish clefts differ to English clefts in that there is more freedom with regards to the type of sentence element that can be fronted (Stenson, 1981). In Irish the structure is as follows: Copula (*is*), followed by the fronted element (Predicate), followed by the rest of the sentence (Relative Clause). The predicate can take

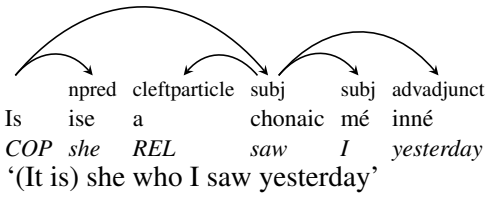

Figure 3: Dependency structure for cleft construction

the form of a pronoun, noun, verbal noun, adverb, adjective, prepositional or adverbial phrase. For example:

- Adverbial Fronting:
  *Is **laistigh de bhliain** a déanfar é*: "It's **within a year** that it will be done"
- Pronoun Fronting:
  *Is **ise** a chonaic mé inné*: "It is **she** who I saw yesterday"

Stenson (1981) describes the cleft construction as being similar to copular identity structures with the order of elements as Copula, Predicate, Subject. This is the basis for the cleft analysis provided by Sulger (2009) in Irish LFG literature. We follow this analysis but with a slight difference in the way we handle the *'a'*. According to Stenson, the *'a'* is a relative particle which forms part of the relative clause. However, there is no surface head noun in the relative clause – it is missing a NP. Stenson refers to these structures as having an 'understood' nominal head such as *an rud* "the thing" or *an té* "the person/the one". e.g. *Is ise **[an té]** a chonaic mé inné*". When the nominal head is present, it becomes a copular identity construction: *She is the one who I saw yesterday*[1]. To distinguish the *'a'* in these cleft sentences from those that occur in relative clauses with surface head nouns, we introduce a new dependency label `cleftparticle` and we attach *'a'* to the verb *chonaic* using this relation. This is shown in Figure 3.

Subject complements In copular constructions, the grammatical subject may take the form of a finite verb clause. In the labelling scheme of Lynn et al. (2012), the verb, being the head of the clause is labelled as a subject (`subj`). We choose to highlight these finite verb clauses as more specific types of grammatical subjects, i.e. subject

[1]Note that this sentence is ambiguous, and can also translate as *She was the one who saw me yesterday*.

complement (csubj)[2]. See Figure 4 for an example.

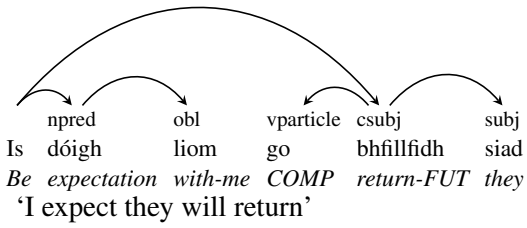

Figure 4: Dependency structure with new subject complement labelling

Wh-questions Notwithstanding Stenson's observation that WH-questions are syntactically similar to cleft sentences, we choose to treat them differently so that their predicate-argument structure is obvious and easily recoverable. Instead of regarding the WH-word as the head (just as the copula is the head in a cleft sentence), we instead regard the verb as the sentential head and mark the WH-element as a dependent of that, labelled as subj_q, obj_q or advadjunct_q. An example of obj_q is in Figure 5.

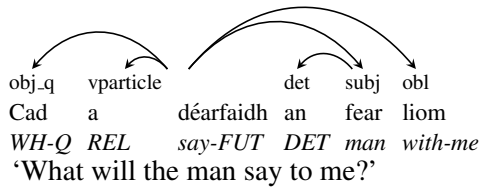

Figure 5: Dependency structure for question construction

2.3 Comparison of Parsing experiments

Lynn et al. (2012) carried out preliminary parsing experiments with MaltParser (Nivre et al., 2006) on their original treebank of 300 sentences. Following the changes we made to the labelling scheme as a result of the second IAA study, we re-ran the same parsing experiments on the newly updated seed set of 300 sentences. We used 10-fold cross-validation on the same feature sets (various combinations of form, lemma, fine-grained POS and coarse-grained POS). The improved results, as shown in the final two columns of Table 3, reflect the value of undertaking an analysis of IAA-1 results.

[2]This label is also used in the English Stanford Dependency Scheme (de Marneffe and Manning, 2008)

3 Active Learning Experiments

Now that the annotation scheme and guide have reached a stable state, we turn our attention to the role of active learning in parser and treebank development. Before describing our preliminary work in this area, we discuss related work .

3.1 Related Work

Active learning is a general technique applicable to many tasks involving machine learning. Two broad approaches are Query By Uncertainty (QBU) (Cohn et al., 1994), where examples about which the learner is least confident are selected for manual annotation; and Query By Committee (QBC) (Seung et al., 1992), where disagreement among a committee of learners is the criterion for selecting examples for annotation. Active learning has been used in a number of areas of NLP such as information extraction (Scheffer et al., 2001), text categorisation (Lewis and Gale, 1994; Hoi et al., 2006) and word sense disambiguation (Chen et al., 2006). Olsson (2009) provides a survey of various approaches to active learning in NLP.

For our work, the most relevant application of active learning to NLP is in parsing, for example, Thompson et al. (1999), Hwa et al. (2003), Osborne and Baldridge (2004) and Reichart and Rappoport (2007). Taking Osborne and Baldridge (2004) as an illustration, the goal of that work was to improve parse selection for HPSG: for all the analyses licensed by the HPSG English Resource Grammar (Baldwin et al., 2004) for a particular sentence, the task is to choose the best one using a log-linear model with features derived from the HPSG structure. The supervised framework requires sentences annotated with parses, which is where active learning can play a role. Osborne and Baldridge (2004) apply both QBU with an ensemble of models, and QBC, and show that this decreases annotation cost, measured both in number of sentences to achieve a particular level of parse selection accuracy, and in a measure of sentence complexity, with respect to random selection.

However, this differs from the task of constructing a resource that is intended to be reused in a number of ways. First, as Baldridge and Osborne (2004) show, when "creating labelled training material (specifically, for them, for HPSG parse se-

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

Table 3: Preliminary MaltParser experiments with the Irish Dependency Treebank: Pre- and post-IAA-2 results

lection) and later reusing it with other models, gains from active learning may be negligible or even negative": the simulation of active learning on an existing treebank under a particular model, with the goal of improving parser accuracy, may not correspond to a useful approach to constructing a treebank. Second, in the actual task of constructing a resource — interlinearized glossed text — Baldridge and Palmer (2009) show that the usefulness of particular example selection techniques in active learning varies with factors such as annotation expertise. They also note the importance of measures that are sensitive to the cost of annotation: the sentences that active learning methods select are often difficult to annotate as well, and may result in no effective savings in time or other measures. To our knowledge, active learning has not yet been applied to the actual construction of a treebank: that is one of our goals.

Further, most active learning work in NLP has used variants of QBU and QBC where instances with the *most* uncertainty or disagreement (respectively) are selected for annotation. Some work by Sokolovska (2011) in the context of phonetisation and named entity recognition has suggested that a distribution over degrees of uncertainty or disagreement may work better: the idea is that examples on which the learners are more certain or in greater agreement might be more straightforwardly added to the training set. This may be a particularly suitable idea in the context of treebank construction, so that examples selected by active learning for annotation are a mix of easier and more complex.

3.2 Setup

The basic treebank/parser bootstrapping algorithm is given in Figure 6. In an initialisation step, a parsing model is trained on a seed set of gold standard trees. In each iterative step, a new batch of unseen sentences is retrieved, the parsing model is used to parse these sentences, a subset of these automatically parsed sentences is selected, the parse trees for the sentences in this subset are manually corrected, the corrected trees are added to the training set and a new parsing model is trained. This process is repeated, ideally until parsing accuracy converges.

$t \leftarrow$ 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

Figure 6: The basic bootstrapping algorithm

We experiment with two versions of this basic bootstrapping algorithm. In the *passive learning* variant, the Y trees that are added to the training data on each iteration are chosen at random from the batch of X unseen sentences. In the *active learning* variant, we select these trees based on a notion of how informative they are, i.e. how much the parser might be improved if it knew how to parse them correctly. We approximate informativeness based on QBC, specifically, disagreement between a committee of two parsers Thus, we rank the set of X trees (u_p) based on their disagreement with a second reference parser.[3] The

[3]This assessment of disagreement between two trees is based on the number of dependency relations they disagree on, which is the fundamental idea of the F-complement measure of Ngai and Yarowsky (2000). Disagreement between

top Y trees from this ordered set are manually revised and added to the training set for the next iteration.

We use MaltParser as the only parser in the passive learning setup and the main parser in the active learning setup. We use another dependency parser Mate (Bohnet, 2010) as our second parser in the active learning setup. Since we have 450 gold trees, we split them into a seed training set of 150 trees, a development set of 150 and a test set of 150. Due to time constraints we run the two versions of the algorithm for four iterations, and on each iteration 50 (Y) parse trees are hand-corrected from a set of 200 (X). This means that the final training set size for both setups is 350 trees (150 + (4*50)). However, the 4*50 training trees added to the seed training set of 150 are not the same for both setups. The set of 200 unseen sentences in each iteration is the same but, crucially, the subsets of 50 chosen for manual correction and added to the training set on each iteration are different — in the active learning setup, QBC is used to choose the subset and in the passive learning setup, the subset is chosen at random. Only one annotator carried out all the manual correction.

3.3 Results

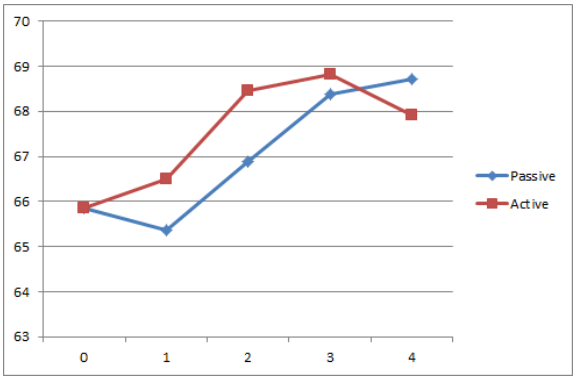

Figure 7: Passive versus Active Learning: **Labelled Attachment Accuracy**. The x-axis represents the number of training iterations and the y-axis the labelled attachment score.

The results of our bootstrapping experiments are shown in Figures 7 and 8. Figure 7 graphs the labelled attachment accuracy for both the passive and active setups over the four training iterations.

two trees, t_1 and t_2 is defined as $1 - LAS(t_1, t_2)$.

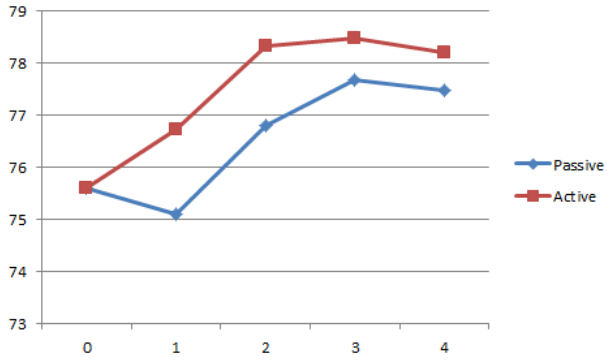

Figure 8: Passive versus Active Learning: **Unlabelled Attachment Accuracy**. The x-axis represents the number of training iterations and the y-axis the unlabelled attachment score.

	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

Table 4: Differences between active and passive training sentences. Correction effort is the level of disagreement between the automatic parse and its correction (1-LAS)

Figure 8 depicts the unlabelled attachment accuracy. All results are on our development set.

3.4 Analysis

On the whole, the results in Figures 7 and 8 confirm that adding training data to our baseline model is useful and that the active learning results are superior to the passive learning results (particularly for unlabelled attachment accuracy). However, the drop in labelled attachment accuracy from the penultimate to the final iteration in the active learning setup is curious.

We measure the difference between the passive and active learning training sentences in terms of sentence length as a way of ascertaining the difference in annotation difficulty between the two sets. Since the training sentences were manually corrected before adding them to the training sets, this means that we can also measure how much correction was involved by measuring the level of disagreement between the automatic parses and their gold-standard corrected versions. This represents another approximation of annotation diffi-

culty.

The results are shown in Table 4. We can see that there is no significant difference in average sentence length between the active and passive learning sets (apart from the final iteration). However, the correction effort figures confirm that the active learning sentences require more correction than the passive learning sentences. This demonstrates that the QBC metric is successful in predicting whether a sentence is hard to parse but it also calls into doubt the benefits of active learning over passive learning, especially when resources are limited. Do the modest gains in parsing accuracy warrant the extra annotation effort involved?

It is interesting that the biggest difference in sentence length is in iteration 4 where there is also a drop in active learning performance on the development set when adding them to the parser. If we examine the 50 trees that are corrected, we find one that has a length of 308 tokens. If this is omitted from the training data, labelled attachment accuracy rises from 67.92 to 69.13 and unlabelled attachment accuracy rises from 78.20 to 78.49. It is risky to conclude too much from just one example but this appears to suggest that if sentences above a certain length are selected by the QBC measure, they should not be revised and added to the training set since the correction process is more likely to be lengthy and error-prone.

The test set shows similar trends to the development set. The baseline model obtains a LAS of 63.4%, the final passive model a LAS of 67.2% and the final active model a LAS of 68.0%, (increasing to 68.1% when the 308-token sentence is removed from the training set). The difference between the active and passive learning results is not, however, statistically significant.

3.5 Making Use of Unlabelled Data

One criticism of the active learning approach to parser/treebank bootstrapping is that it can result in a set of trees which is an unrepresentative sample of the language since it is skewed in favour of the type of sentences chosen by the active learning informative measure. One possible way to mitigate this is to add automatically labelled data in addition to hand-corrected data. Taking the third active learning iteration with a training set of 300 sentences as our starting point, we add automatic parses from the remaining sentences in the unlabelled set for that iteration. The unlabelled set is ordered by disagreement with the reference parser and so we keep adding from the bottom of this set until we reach the subset of 50 trees which were manually corrected, i.e. we prioritise those parses that show the highest agreement with the reference parser first because we assume these to be more accurate. The results, shown in Figure 9, demonstrate that the addition of the automatic parses makes little difference to the parsing accuracy. This is not necessarily a negative result since it demonstrates that the training sentence bias can be adjusted without additional annotation effort and without adversely affecting parsing accuracy (at least with this limited training set size).

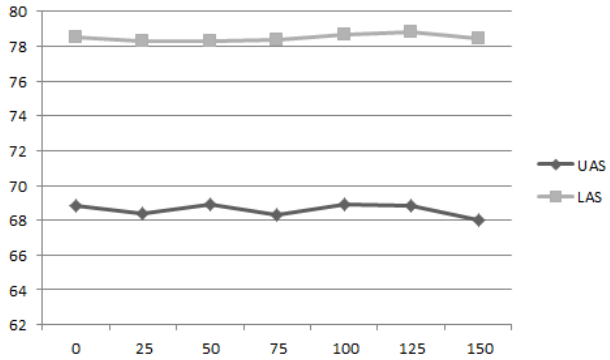

Figure 9: Adding Automatically Parsed Data to the Training set: the x-axis shows the number of automatically parsed trees that are added to the training set and the y-axis shows the unlabelled and labelled attachment accuracy on the development set.

4 Conclusion

We have presented the finalised annotation scheme for the Irish Dependency Treebank and shown how we arrived at this using inter-annotator agreement experiments, analysis and discussion. We also presented the results of preliminary parsing experiments exploring the use of active learning. Future work involves determining the length threshold above which manual annotation should be avoided during bootstrapping, experimenting with more active learning configurations, and, of course, further manual annotation.

5 Acknowledgements

The authors would like to thank Josef van Genabith and the three anonymous reviewers for their insightful comments and suggestions.

References

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, December.

Jason Baldridge and Miles Osborne. 2004. Active Learning and the Total Cost of Annotation . In *Proceedings of EMNLP 2004*, pages 9–16, Barcelona, Spain.

Jason Baldridge and Alexis Palmer. 2009. How well does active learning *actually* work? Time-based evaluation of cost-reduction strategies for language documentation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 296–305, Singapore.

Timothy Baldwin, Emily M. Bender, Dan Flickinger, Ara Kim, and Stephan Oepen. 2004. Road testing the English Resource Grammar over the British National Corpus. In *Proceedings of LREC*.

Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *In Proceedings of COLING*.

Joan Bresnan. 2001. *Lexical Functional Syntax*. Oxford: Blackwell.

Özlem Çetinoğlu, Jennifer Foster, Joakim Nivre, Deirdre Hogan, Aoife Cahill, and Josef van Genabith. 2010. LFG without c-structures. In *Proceedings of the 9th International Workshop on Treebanks and Linguistic Theories (TLT9)*.

Jinying Chen, Andrew Schein, Lyle Ungar, and Martha Palmer. 2006. An Empirical Study of the Behavior of Active Learning for Word Sense Disambiguation. In *Proceedings of the Human Language Technology Conference of the NAACL*, pages 120–127, New York City, USA, June.

David Cohn, Les Atlas, and Richard Ladner. 1994. Improving generalization with active learning. *Machine Learning*, 15(2):201–221.

Mary Dalrymple, Helge Dyvik, and Tracy Holloway King. 2004. Copular complements: Closed or open? In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '04 Conference*.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *Workshop on Crossframework and Cross-domain Parser Evaluation (COLING2008)*.

Sašo Džeroski, Tomaž Erjavec, Nina Ledinek, Petr Pajas, Zdenek Žabokrtsky, and Andreja Žele. 2006. Towards a Slovene dependency treebank. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Katri Haverinen, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Filip Ginter, and Tapio Salakoski. 2010. Treebanking Finnish. In *Proceedings of The Ninth International Workshop on Treebanks and Linguistic Theories (TLT9)*, pages 79–90.

Steven C. H. Hoi, Rong Jin, and Michael Lyu. 2006. Large-scale text categorization by batch mode active learning. In *Proceedings of the 15th International World Wide Web Conference (WWW 2006)*, pages 633–642, Edinburgh, UK.

Rebecca Hwa, Miles Osborne, Anoop Sarkar, and Mark Steedman. 2003. Corrected co-training for statistical parsers. In *Proceedings of the Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, Washington DC, US.

John Judge, Aoife Cahill, and Josef van Genabith. 2006. Questionbank: Creating a corpus of parse-annotated questions. In *In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*.

Adam Kilgarriff, Michael Rundell, and Elain Uí Dhonnchadha. 2006. Efficient corpus creation for lexicography. *Language Resources and Evaluation*, 18(2).

Matthias Kromann. 2003. The Danish Dependency Treebank and the DTAG Treebank Tool. In *Proceedings from the 2nd Workshop on Treebanks and Linguistic Theories (TLT 2003)*.

David Lewis and William Gale. 1994. A Sequential Algorithm for Training Text Classifiers. In *Proceedings of the 17th Annual International ACL-SIGIR Conference on Research and Development of Information Retrieval*, pages 3–12, Dublin, Ireland.

Teresa Lynn, Ozlem Cetinoglu, Jennifer Foster, Elaine Uí Dhonnchadha, Mark Dras, and Josef van Genabith. 2012. Irish treebanking and parsing: A preliminary evaluation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*.

Grace Ngai and David Yarowsky. 2000. Rule writing or annotation: Cost-efficient resource usage for base noun phrase chunking. In *Proceedings of ACL*.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Fredrik Olsson. 2009. A literature survey of active machine learning in the context of natural language processing. Technical Report T2009:06, SICS.

Miles Osborne and Jason Baldridge. 2004. Ensemble-based Active Learning for Parse Selection. In *HLT-NAACL 2004: Main Proceedings*, pages 89–96, Boston, MA, USA.

Roi Reichart and Ari Rappoport. 2007. An Ensemble Method for Selection of High Quality Parses. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 408–415, Prague, Czech Republic.

Tobias Scheffer, Christian Decomain, and Stefan Wrobel. 2001. Active hidden Markov models for in-

formation extraction. In *Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis (IDA-2001)*, pages 309–318.

Sebastian Seung, Manfred Opper, and Haim Sompolinsky. 1992. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 287–295, Pittsburgh, PA, US.

Nataliya Sokolovska. 2011. Aspects of Semi-Supervised and Active Learning in Conditional Random Fields. In *Proceedings of the European Conference on Machine Learning (ECML PKDD) 2011*, pages 273–288.

Nancy Stenson. 1981. *Studies in Irish Syntax*. Gunter Narr Verlag Tübingen.

Sebastian Sulger. 2009. Irish clefting and information-structure. In *Proceedings of the LFG '09 Conference*.

Cynthia Thompson, Mary Elaine Califf, and Raymond Mooney. 1999. Active learning for natural language parsing and information extraction. In *Proceedings of the Sixteenth International Machine Learning Conference (ICML-99)*, pages 406–414, Bled, Slovenia.

Elaine Uí Dhonnchadha, Caoilfhionn Nic Pháidín, and Josef van Genabith. 2003. Design, implementation and evaluation of an inflectional morphology finite state transducer for Irish. *Machine Translation*, 18:173–193.

Elaine Uí Dhonnchadha. 2009. *Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar*. Ph.D. thesis, Dublin City University.