Morphological Features of the Irish Universal Dependency Treebank

Teresa Lynn[1], Jennifer Foster[1] and Mark Dras[2]
[1]ADAPT Centre, School of Computing, Dublin City University, Ireland
[2]Department of Computing, Macquarie University, Sydney, Australia
[1]{teresa.lynn, jennifer.foster}@adaptcentre.ie
[2]{mark.dras}@mq.edu.au,

Abstract

The Universal Dependencies Project[1] (Nivre, [9]; Nivre et al., [10]) is an ongoing effort towards creating a set of harmonised dependency treebanks that are annotated and structured according to universal guidelines. This paper reports on the addition of morphological features to the Irish Universal Dependencies Treebank (IUDT). Our feature set subscribes to the feature inventory of the UD Project and has been mapped from Irish morpho-syntactic tags – the output of a Finite State Morphological Analyser for Irish (Uí Dhonnchadha and van Genabith [16]). Irish, a Celtic language, has some relatively unusual morphological features that require language-specific labels not covered by the universal feature set. In this paper, we summarise the Irish-specific features that we have added to this set by explaining the linguistic properties that they each describe. We also report on the first parsing experiments using the IUDT by assessing the effect that the inclusion of morphological features has on parsing accuracy.

1 Introduction

The motivation behind the Universal Dependencies Project (Nivre, [9]; Nivre et al., [10]) is to create a set of harmonised dependency treebanks that will facilitate improved multilingual parsing and better cross-linguistic analysis. Treebank development for any language is both resource- and time-intensive. Building a large-scale, fully-annotated treebank, usually requires a large team of linguists and/or computational linguists, all of whom are collectively responsible for both the design and annotation of the dataset. Low-resource languages, however, lack the financial investment enjoyed by those better resourced and more widely researched languages, and as such, resources such as treebanks often take much longer to produce.

[1]http://universaldependencies.org

Irish is a relatively low-resourced language amongst the UD treebank collection. The inclusion of low-resourced languages in large projects like this means that the language can benefit from the experience and contributions of the wider research community. The Irish Universal Dependency Treebank (IUDT) was developed as a result of a mapping of the Irish Dependency Treebank (IDT) annotation scheme (Lynn [6]) to the UD annotation scheme (Lynn and Foster [5]). Both treebanks are currently relatively small in size (1020 trees) and lack morphological features in their initial release. The natural next step in development is therefore the inclusion of such features.

In this paper, we discuss our extension of the IUDT to include morphological features. We follow the UD guidelines[2] for feature inclusion. While the UD project defines a feature set that aims to cover linguistic universals shared across the numerous[3] languages that are part of the project, it is also possible for each language group to define their own language-specific features where necessary. For example, one of the Irish-specific UD features relates to initial mutation – a linguistic phenomenon that is common across Celtic languages (see Section 4.1). However, initial mutation does not feature widely in other languages, and is therefore not deemed 'universal'. Here, we list the morphological tags from the UD feature set that are relevant to Irish and additional new Irish-specific features.

Additionally, we consider that parsing models for morphologically rich languages can benefit from the inclusion of morpho-syntactic features given sufficient training data. This additional level of annotation in dependency treebanks, at the morpho-syntactic level, contributes to a deeper understanding of the data that may not be achieved through part-of-speech (POS)-tags or dependency labels alone. This has been demonstrated in various studies, including shared tasks on the parsing of morphologically rich languages (e.g. Seddah et al, [12]; [13]). To this end, we evaluate the inclusion of these features in the IUDT through empirical methods.

2 The morphology of Irish

In computational terms, Irish is regarded as a morphologically rich language (Lynn et al [7]), and as such, Irish parsing models should benefit from the inclusion of morphological features in the training data. Stenson [14] describes the Irish language as an "inflectional language, tending more toward isolating than polysynthetic in general". She also notes that inflection is primarily realised through suffixation, yet initial mutation (a characteristic of Celtic languages) is also common and appears in the form of eclipsis (e.g. *bord*; *ar an* ***m****bord* 'on the table') or lenition (e.g. *dath an b****h****oird* 'the colour of the table'). Another prominent feature (and also of Scottish Gaelic and Manx), which influences inflection, is the existence of two sets of consonants, referred to as 'broad' and 'slender' consonants (Ó Siadhail [11]). For example, *buail* 'hit' + *adh* (verbal noun suffix) → *bu****a****l****a****dh*. The fol-

[2] http://universaldependencies.org/u/feat/index.html

[3] The latest release of UD treebanks (v.4.) includes 64 treebanks, covering 47 languages.

lowing is a short summary of the main inflectional processes in Irish. For a more detailed description, see The Christian Brothers [3] or Lynn [6].

Nouns In terms of nominal inflection, Modern Irish only uses three cases: common, genitive and vocative. The common case covers nominative, accusative and dative, yet it should be noted that the dative case is marked on personal pronouns. Each noun falls into one of five declensions. Case, declension and gender are all expressed through inflection.

Verbs Verbs can be marked for both tense and aspect, and inflect for person and number (e.g. *ith* 'eat'; ***d'ithimis*** '**we used to** eat').

Adjectives The Christian Brothers ([3] p.63) note eight main declensions of adjectives. They can inflect for genitive singular masculine, genitive singular feminine and nominative plural (e.g. *bacach* 'lame'; *baca**igh*** Gen.Sg.Masc).

Prepositions Simple prepositions can inflect for a pronominal object, indicating person and number (e.g. *le* 'with'; *liom* 'with me'; *lei* 'with her').

3 Irish Morphosyntactic Tagset

The IDT and IUDT were built upon a gold-standard POS-tagged corpus of Modern Irish, which was developed by Uí Dhonnchadha [15]. The POS tags in this corpus are the output format of the Irish Morphological Analyser (Uí Dhonnchadha and van Genabith [16]). They are based on a mapping from the Irish PAROLE Morphosyntactic Tagset (ITÉ [4]) to a Finite State Morphological Feature Tagset.

The following is an example sentence labelled with FST tagset output: *Tá sé soiléir* 'It is clear'. Each token is followed by a string of POS tags and morphosyntactic features, separated by '+'. The string `sé sé+Pron+Pers+3P+Sg+Masc` tells us that `sé` (with the same lemma form) is a 3rd person singular personal pronoun, of masculine gender.

(1) *Tá sé soiléir* 'It is clear'.

```
Tá bí+Verb+PresInd
sé sé+Pron+Pers+3P+Sg+Masc
soiléir soiléir+Adj+Base
```

Until now, the IUDT contained only the POS tags (coarse- and fine-grained) of this feature set, i.e. `+Pron+Pers`, indicating personal pronoun. Our purpose of now including additional morphosyntactic features in treebank data is to mark additional lexical and grammatical properties of words that are not available through POS tags alone. As shown in (1) above, such additional morphological features for

Irish data are readily available for inclusion in the data set, but need to be converted to the correct representation and align with the UD guidelines.

The IUDT is in the CoNLL format (Buchhloz [2]), where morphological features are labelled in the FEATS column, and every feature has the form Feature-Name=Value. Every word can have any number of features, separated by the vertical bar (i.e. $a = x|b = y|c = z$). The features for the pronoun (*sé*) in (1) above would therefore be represented as `Gender=Masc|Number=Sing|Person=3`. According to UD guidelines, multiple features (and where applicable, multiple values) should be ordered in alphabetical order.[4]

4 Mapping Irish Morphosyntactic Tagset to the UD feature set

The Universal Dependencies feature set is a standardised list of morphological features, which is based on the Interset system (Zeman [17]). It is divided into lexical and inflectional features. Mapping to the UD feature set was straightforward for the majority of the tags in the Irish Morphosyntactic Tagset. However, we have introduced a number of non-universal features and values in order to fully represent specific linguistic phenomena in the Irish language (see Section 4.1). Table 1 presents the full inventory. Rather than discuss the details of entire feature set here, we only discuss in detail the newly introduced Irish language-specific features (see bolded features). For information on all other standard (universal) features, we point the reader to the Universal Dependencies website.[5]

4.1 Description of Irish-specific features

There are a number of morphosyntactic features in Irish for which the UD universal feature set does not fully account. This is not surprising, as while all languages share certain linguistic universals, what makes languages unique is often seen in their differences across syntax or morphology. For a more fine-grained annotation, we extend the UD feature set to cover these nuances in Irish. Here, we provide a short description of each of the Irish-specific features.[6]

Dialect=Munster, Connaught, Ulster There are three main dialects of Irish and there are a number of lexical items that differ across them. Some surface forms differ, while sharing the same lemma. For example, the past tense form of the lemma *cuir* 'put' is *chuireas* in the Munster dialect, but *chuir* in the Connaught and Ulster dialects. Other lexical items differ in both surface and lemma form. For

[4] http://universaldependencies.org/u/overview/morphology.html

[5] http://universaldependencies.org/u/feat/all.html

[6] Not requiring much discussion, note that the Feature Value `NomAcc` will replace the incorrectly used `Com` tag in future versions of the IUDT to indicate for the common case in Irish. `Com` is used in other UD treebanks to indicate the Comitative case.

Feature Name	Feature Value	Lexical or Inflectional?	Prevalence
Case	Dat, Gen, NomAcc†, Voc	Inflectional	6948
Degree	Cmp, Pos, Sup	Inflectional	595
Dialect†	Connaught, Munster, Ulster	Inflectional	51
Definite	Def, Ind	Inflectional	2553
Form†	Ecl, Emph, hPref, Len, VF	Inflectional	4566
Gender	Fem, Masc	Inflectional	7920
Mood	Cnd, Imp, Ind, Int†, Sub	Inflectional	1917
Negative	Neg	Inflectional	359
NounType†	NotSlender, Slender, Strong, Weak	Inflectional	239
NumType	Card, Ord, Pers	Lexical	165
Number	Plur, Sing	Inflectional	10,529
PartType†	Ad, Cmpl, Comp, Cop, Deg, Inf, Num, Pat, Vb, Voc	Lexical	1526
Person	1, 2, 3	Inflectional	1831
Poss	Yes	Lexical	444
PrepForm†	Cmpd	Lexical	166
PronType	Art, Dem, Ind, Int, Prs, Rel	Lexical	3259
Reflex	Yes	Lexical	61
Tense	Fut, Past, Pres	Inflectional	2163
VerbForm	Cop†, Ger, Inf, Part	Inflectional	1328
Voice	Auto†	Inflectional	251

Table 1: Irish-specific Features: Each `Feature Name` has a list of possible `Feature Values`. `Prevelance` indicates the number of times a feature name appears in the entire treebank of 23,684 tokens. Note that some tokens contain features with multiple values. † indicates newly introduced Irish-specific Feature Names and Feature Values.

example, the word *achan* 'every' is used in Ulster dialect, but *gach* 'every' is used in the Munster and Connaught dialects.

Form=Ecl, Emph, Len, hPref Initial mutation is a feature of Celtic languages. It is triggered by a preceding word and affects the spelling of nouns, adjectives and verbs. Eclipsis (`Ecl`) on verbs occurs following clitics such as interrogative particles (an, nach); complementisers (go, nach); and relativisers (a, nach) (Stenson [14], pp. 21-26). For example, *tugann sé* 'he gives'; *an **d**tugann sé.* 'does he give?' An example of lenition (`Len`) on proper nouns is following the vocative particle *a*. E.g. *Máire* 'Mary'; *A **Mh**áire!* In addition, some words can trigger the h-prefix (`hPref`) on the following word. For example, *le **h**institiúidí* 'with institutes'. Finally, (`Emph`) notes the emphatic marker in Irish. For example, *liom* 'with me'; *liom**sa*** 'with **me**'.

Form=VF VF (Vowel Form) is an indicator of spelling changes that occur in copular verbs when followed by a word that begins with a vowel or a lenited consonant. For example, ***Is*** *féidir liom* 'I can'; *Conas* ***ab fh****éidir liom* 'How can I?'.

Mood=Int While regular verbs make use of interrogative particles (e.g. ***ar*** *chuala tú?* '**did** you hear?') to indicate a question construction, the copula has a number of different forms to indicate the interrogative mood. For example *Is maith leat* 'you like'; ***an*** *maith leat?* 'do you like?'; ***ar*** *mhaith leat?* 'would you like?'; ***nach*** *mhaith leat?* 'do you not like?'; ***nár*** *mhaith leat?* 'would you not like?'

NounType=Weak, Strong Plural nouns are either referred to as Weak Plurals or Strong Plurals. The form of a Strong Plural remains unchanged, regardless of grammatical case. In other words it does not inflect. For example: *Tá na ríthe ag troid* 'the kings are fighting' (Com.Pl); *bás na* ***ríthe*** 'death of the kings' (Gen.Pl). On the other hand, the plural form of a noun is weak if it meets specific criteria related to the common plural form (The Christian Brothers [3] p.34). The weak/strong nature of nouns in turn affects the declension of their modifying adjectives. Therefore, this feature applies to both Nouns and Adjectives.

NounType=NotSlender, Slender Irish has broad (`NotSlender`) and slender consonants. This feature is influenced by a consonant's preceding vowel: a broad consonant is preceded by a broad vowel and a slender consonant by a slender vowel. In some cases, this consonantal feature can impact the spelling of subsequent adjectives. For example, if a plural noun has a slender ending, the following adjective is lenited (e.g. *deonach* 'big'; *eagra****i****s d****h****eonacha* 'voluntary organisations'). This feature applies to Adjectives.

PartType=Ad, Cmpl, Comp, Cop, Deg, Inf, Num, Pat, Vb, Voc Irish makes use of a broad range of particles. It is important to differentiate these particles, because in some cases they share the same form, yet have different functions. For example, *a* can be a vocative particle, relative particle, infinitive particle or quantifier particle, and trigger a variety of different spelling changes (e.g. eclipsis, h-prefix or lenition).

PrepForm=Cmp Compound prepositions contain a simple preposition and a noun. Nouns that follow compound prepositions are inflected in the genitive case. For example, *ar fud* 'throughout' and *domhan* 'world', when combined, becomes: ***ar fud*** *an domha****i****n* 'throughout the world'.

VerbForm=Cop Irish has two verbs 'to be' – the copular verb and the substantive verb (Stenson [14] p.92). The copula is used for (i) classification ('he is a man'), (ii) identification ('Mary is the doctor'), (iii) to express ownership (iv) to mark emphasis through clefting and (v) in making comparisons. The substantive is used

in all other cases. The copula does not inflect for mood, gender or number in the way regular verbs do. As the UD POS tagset subsumes the copula as `VERB`, it is important to distinguish between them through their morphological features, as they follow a different argument structure. The substantive verb follows the general VSO (Verb Subject Object) order, whereas the copula construction mostly follows a Copula-Predicate-Subject order.

Voice=Auto Irish does not have an equivalent to the English passive construction (The Christian Brothers (1988, p.120) and Stenson [14] p.145). Stenson identifies `autonomous verbs` as one type of construction that is used instead. Autonomous verbs have an 'understood' subject and therefore the noun which follows is usually the direct object. For example, *chonacthas iad* 'they were seen' translates literally as 'someone saw them'. The dative case of the pronoun *iad* 'them' (vs *siad* 'they') clearly marks it as the object of the verb.

5 Experiments

Treebank	**Features**	**LAS (test)**	**UAS (test)**	**LAS (dev)**	**UAS (dev)**
IUDT (baseline)	no	68.27	76.29	70.59	78.73
IUDT	yes	70.63	77.57	71.79	79.45
IUDT + Opt	yes	**72.18**	**78.3**	**73.33**	**79.52**
IUDT	universal only	69.26	76.99	70.70	78.66
IDT	no	69	78.33	70.66	79.8
		LAS (10-fold)	**UAS (10-fold)**		
IUDT 10-fold	yes	72.19	79.12		

Table 2: Parsing experiment results (LAS = Labelled Attachment Score; UAS = Unlabelled Attachment Score; Opt = using MaltOptimizer)

In this section we discuss a number of parsing experiments carried out on the IUDT. We use MaltParser (Nivre et al., [8]) for all our experiments.[7] We first look at how the inclusion of morphological features impacts parsing models trained on the IUDT. We then follow on to optimise the inclusion of these features through the use of MaltOptimizer. Following that, we demonstrate the impact of specifying the Irish-specific features in our set. Then, we put our parsing results into context by providing parsing results for the IDT. Finally, we provide parsing results based on 10-fold cross-validation in order to provide a more realistic parsing accuracy report.

5.1 Evaluation of inclusion of morphological features

In order to assess the benefit of including morphological features in the IUDT, we report here on parsing experiments that are based on models trained on two

[7] MaltParser v1.7, available to download from http://www.maltparser.org/download.html

different versions of the treebank:

- our baseline, which is the original IUDT (v1.0)[8], (Lynn and Foster [5])
  – *without* morphological features
- the newly updated IUDT (v1.4)[9]
  – *includes* the morphosyntactic information (Section 2).

In this experiment, we follow the development/test/training treebank split, as per the UD convention, i.e. roughly 10%-10%-80%: dev (150 trees)/ test (150 trees)/ training (remaining = 720 trees). We can see from the results presented in Table 2 that the inclusion of the morphological features in the IUDT leads to a clear increase in parsing accuracy. As the only difference across the two experiments is the presence or absence of morphological data (e.g. there is no change in annotation scheme, standard treebank data or MaltParser settings), we can conclude that the morphological information has a positive impact on parsing accuracy.

5.2 Optimising feature inclusion

With the promising results we get from including morphological information in the training data, we are prompted to evaluate the use of MaltOptimizer (Ballesteros [1]) as part of our efforts to fully optimise the inclusion of such rich features. MaltOptimizer is a freely available tool that can be used in conjunction with MaltParser to optimise parsing, based on an analysis of training data. The analysis phase collects information such as the number of words/sentences, percentage of projective/non-projective trees, existence of covered roots[10], features in LEMMA, FEATS, CPOS (coarse-grained part-of-speech) and POS (fine-grained part-of-speech) columns. Based on this, the output provides the user with the most effective combination of options (user-defined parameters) to use in the MaltParser training and parsing phases.

MaltOptimizer defined the optimal training options for the IUDT dataset as the Nivre arc-eager parsing algorithm with normal root handling and a LIBLINEAR learner classifier. As we show in Table 2, this configuration resulted in a LAS increase of 1.55 and UAS increase of 0.73 on the test set. The improvement on the results of the dev set were similar for LAS (1.54), but less so for UAS (0.07).

5.3 Impact of Irish-specific morphological features

In Section 4.1, we list and describe the Irish language-specific features that we have introduced to our feature set. We acknowledge that we are motivated to specifying these new features as we *feel* that this linguistic phenomena is integral to the syntax of the Irish language and should be captured in the data. This intuition comes

[8] http://hdl.handle.net/11234/1-1464

[9] https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1827

[10] Covered roots are cases where the root (head value = 0) is crossed by one or more arcs.

from a knowledge of the language, yet it is also important to empirically justify this inclusion. For this reason, we experiment on a version of the IUDT treebank that includes only the universal features, and excludes the Irish-specific features. Our results are also presented in Table 2, where we can see that removing the Irish specific features results in a drop in parsing accuracy, thus supporting the motivation for their general inclusion in the latest version.

On closer analysis of the parser output (development set), we observe exactly how some of these features help to improve parsing. For example, the feature `Voice=Auto` – which we explain in Section 4.1 indicates the autonomous verb – appears to help the parser in correctly identifying the direct object (`dobj`) argument of the verb. Without this feature, confusion arises when the parser assumes that the noun following the autonomous verb is the subject, thus assigning the `nsubj` label attachment. In some instances, correct identification of the verb–direct object attachment resulted in a knock-on effect of improved parsing of the overall tree. Likewise, we observe that the morphological feature `PrepForm=Cmpd` has helped the parser to more accurately identify the correct dependency attachment and labelling of `case` to compound prepositions, where previously parser confusion led to the incorrect assignment of either `root` or `nmod` labels. Again, improvements in this labelling had a knock-on effect in some cases, thus improving the overall parse of the tree.

5.4 Comparison with IDT parsing results

To put the parsing results for Irish Universal Dependencies into context, we look at comparing our IUDT parsing results with those of the original Irish Dependency Treebank (IDT). It should be noted that the IDT does not yet contain morphological features. The results are presented towards the bottom of Table 2.

A number of factors can affect the quality of parsing, such as the design of an annotation scheme, the number of dependency labels that are used (including granularity of labels) and the settings used in a parsing framework. As reported by Lynn and Foster [5], the IDT has 47 labels, compared to just 35 used in the IUDT. In addition, due to the different annotation schemes, both the label names and structural analyses differ across treebanks. It is therefore unsurprising that the parsing results based on the IDT differ from those based on the IUDT. We remark here that accuracy *dropped* across the board for the parsing model trained on the IUDT. It is interesting to note that a scheme that is designed specifically for Irish appears to be easier to parse than UD, which instead aims to be universal, and is informed by a large collection of languages.

We also note that adding morphological features to the IUDT increased parser accuracy to a level which is on a par with the IDT parser (without morphological features). This could suggest that the inclusion of morphological features overcomes the potential loss of information introduced when mapping treebank annotations from an Irish-language inspired scheme to a universal scheme. However, it should be noted that strong conclusions cannot be drawn from this, as the Malt-

Parser settings chosen for these parsing experiments were optimised and tuned to the IDT and may not necessarily be optimal for the IUDT.[11]

5.5 Cross-validation

It is worth noting that the baseline results provided here for the IDT are actually lower than those previously reported by Lynn (2016) (LAS 71.4% and UAS 80.1%). The reason for this is the difference in the data splitting approach for the dev/test/ train sets. As mentioned in Section 5.1, our IDT (and IUDT) results presented here are based on the data split convention of the UD project. Prior reported IDT experiments, however, have been carried out using k-fold cross-validation.

While a conventional data-split approach across a large project is well motivated, parsers trained and tested on small datasets may not always reflect the full truth of attainable parsing accuracy. Instead, parsing model evaluation is often carried out using k-fold cross-validation when working with smaller data sets. This approach involves splitting the data into k sets of equal size, one set used as a test set and the remaining as training data, and repeated iteratively until all partitions have served as a test set once. The result is an average across k test set results.

Thus, in order to acquire a more realistic set of results in our current experiments, we perform 10-fold cross-validation on the IUDT with morphological features to get a clearer idea of parsing quality. Our results show an increase across both LAS (70.63% – 72.19%) and UAS (77.57% – 79.12%).

6 Conclusion

We have described the morphological features recently introduced to the Irish Universal Dependency treebank, in the form of both standard UD features and features introduced specifically for the Irish language. Parsing experiments with MaltParser suggest that these morphological annotations are helpful, and results can be further improved using MaltOptimizer. Ablation experiments demonstrate that the Irish-specific morphological information has a useful role to play.

7 Acknowledgements

We would like to thank the anonymous reviewers for their valuable feedback and Lamia Tounsi for her analysis advice. This work has been supported by the ADAPT Centre for Digital Content Technology (www.adaptcentre.ie) at Dublin City University, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

[11] The MaltParser configurations on previously reported Irish parsing experiments included stacklazy and liblinear learner algorithms. The feature set included word form, lemma and both coarse and fine-grained POS tags.

References

[1] Ballesteros, Miguel (2012) MaltOptimizer: A System for MaltParser Optimization, In *Proceedings of the Eighth International Conference on Linguistic Resources and Evaluation (LREC)*, Istanbul, 2012.

[2] Buchhloz, Sabine and Marsi,Erwin (2006)CoNLL-X shared task on Multilingual Dependency Parsing, In Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL), New York City, USA, 2006.

[3] The Christian Brothers (1998) *New Irish Grammar* Dublin: CJ Fallon

[4] ITÉ (2002) PAROLE Morphosyntactic Tagset for Irish. Institiúid Teangeolaíochta Éireann.

[5] Lynn, Teresa and Foster, Jennifer (2016) Universal Dependencies for Irish. In *Proceedings of the Second Celtic Language Technology Workshop*, Paris, France, 2016.

[6] Lynn, Teresa (2016) *Irish Dependency Treebanking and Parsing*. PhD Thesis, Dublin City University and Macquarie University Sydney, 2016.

[7] Lynn, Teresa, Foster, Jennifer, Dras, Mark (2013) Working with a small dataset – semi-supervised dependency parsing for Irish. In *Proceedings of Statistical Parsing of Morphologically Rich Languages*, Seattle, USA, 2013.

[8] Nivre, Joakim, Hall, Johan, Nilsson, Jens (2006) MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy, 2006

[9] Nivre, Joakim (2015) Towards a Universal Grammar for Natural Language Processing. In *Computational Linguistics and Intelligent Text Processing*, Springer International Publishing, 2015.

[10] Nivre, Joakim, de Marneffe, Marie-Catherine, Ginter, Filip, Goldberg, Yoav, Hajic, Jan, Manning, Christopher D., McDonald, Ryan, Petrov, Petrov, Pyysalo, Sampo, Silveira, Natalia, Tsarfaty, Reut, Zeman, Daniel (2016). Universal Dependencies v1: A Multilingual Treebank Collection. In Proceedings of the Tenth International Conference on Language Resources and Evaluation. Portoroz, Slovenia, 2016.

[11] Ó Siadhail, Mícheál (1989), *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

[12] Seddah, Djamé, Kubler, Sandra, Reut, Tsarfaty (2014) Introducing the SPMRL 2014 Shared Task on Parsing Morphologically-Rich Languages. In *Proceedings of the first joint meeting of Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical English*, Dublin, Ireland, 2014.

[13] Seddah, Djame and Tsarfaty, Reut and Kuebler, Sandra and Candito, Marie and Choi, Jinho and Farkas , Richard and Foster, Jennifer and Goenaga, Iakes and Gojenola, Koldo and Goldberg, Yoav and Green, Spence and Habash, Nizar and Kuhlmann, Marco and Maier, Wolfgang and Nivre, Joakim and Przepiórkowski, Adam and Roth, Ryan and Seeker, Wolfgang and Versley, Yannick and Vincze, Veronika and Wolinski, Marcin and Wróblewska, Alina and Villemonte de la Clérgerie, Eric (2013) Overview of the SPMRL 2013 shared task: cross-framework evaluation of parsing morphologically rich languages. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically Rich Languages*, Seattle, WA, 2013.

[14] Stenson, Nancy (1981) *Studies in Irish Syntax*. Tübingen: Gunter Narr Verlag

[15] Uí Dhonnchadha, Elaine (2009) *Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar*, PhD Thesis, Dublin City University, 2009.

[16] Uí Dhonnchadha, Elaine, van Genabith, Josef (2006) A Part-of-speech tagger for Irish using Finite-State Morphology and Constraint Grammar Disambiguation. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 2006.

[17] Zeman, Dan (2008) Reusable Tagset Conversion Using Tagset Drivers. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.