

A Striking Takes Place: A Lexical-Functional Analysis of the Irish Autonomous Verb

James H. West

University of Georgia – jhw25911@uga.edu

Abstract

In this paper, a unified analysis of the Irish autonomous verb across several attested constructions is presented using the formalism of Lexical-Functional Grammar. Namely, the present analysis proposes that no semantic argument is assigned to the subject of the autonomous verb, which unifies its various applications in intransitive, transitive, and raising-like constructions. This is followed by a computational application of the theoretical analysis using the Xerox Linguistic Environment. A computer grammar was written for this application that is capable of parsing novel sentences within the distribution of its ruleset and lexicon. Data and graphical analyses from this computational application are presented, followed by a discussion of the applications of this analysis.

1 Introduction

This paper examines the syntactic puzzle posed by a verb form known as the autonomous in the Irish language. Often translated into English using a passive construction (1 a), the Irish autonomous form is actually fully productive for all verbs in the language, including intransitives, but in contrast to this productivity stands a lexically-restricted phenomenon using the autonomous form of particular verbs to create a raising-like construction. This latter phenomenon thus poses the challenge of determining a unified analysis for the autonomous form as a whole.

The autonomous verb stands apart from typical passive constructions by keeping its object in the accusative case, shown by the unacceptability of (1 b), which resembles the cross-linguistic concept of passive constructions and promotes its object to a nominative subject. The autonomous form surpasses typical passive constructions in productivity, which means it can be applied to intransitive verbs to form a complete and grammatical sentence, as in (1 c). Brothers (1902) noted that even transitive verbs like *buail* ‘to strike’ can be used in the autonomous form without an object, translating the sentence *Buail-tear.* as ‘The action of striking takes place’.

- (1) (a) *Feictear* *é*.
 see-PRES-AUTO PRO-3SG-MASC-ACC
 ‘He is seen.’
 (lit. ‘*Him is seen.’)
- (b) **Feictear* *sé*.
 see-PRES-AUTO PRO-3SG-MASC-NOM
 ‘He is seen.’
- (c) *Siúladh*.
 walk-PAST-AUTO
 ‘(Someone) walked.’

Compare also the autonomous construction in (2a) to the coexistent construction in (2b) that does perfectly match the cross-linguistic expectation for the passive, and employs a verbal adjective, similar to the equivalent English sentence. Furthermore, note the morphological resemblance of the autonomous verb to a pronominal subject in Irish (compare (2c) and (2d)), but that it is restricted from participating in subject-shifting constructions available to pronouns that express semantic agents, such as (2e). It is impossible to form a grammatical equivalent to this latter example using the Irish autonomous.

- (2) (a) *Léitear* *an leabhar*.
 read-PRES-AUTO the book
 ‘The book is read.’
 (or ‘Someone reads the book.’)
- (b) *Tá* *sé* *léite*.
 be-PRES PRO-3SG-MASC-NOM read-VAdj
 ‘It is read.’
- (c) *Buailim* *an liathróid*.
 hit-PRES-1SG the ball
 ‘I hit the ball.’
- (d) *Buailtear* *an liathróid*.
 hit-PRES-AUTO the ball
 ‘The ball is hit.’
 (or ‘(Someone) hits the ball.’)
- (e) *Tá* *leabhar agam*.
 be-PRES a book at PRO-1SG
 ‘I have a book.’
 (lit. ‘A book is at me.’)

This paper argues that the subject of the autonomous verb does not receive a semantic role assignment, and merely serves to fill the “subject” slot in the syntax in sentences where no subject is specified. In this way, it resembles impersonal

verbs with expletive-*it* constructions in English such as “*It* rains.” or “*It* seems that...”, where *it* does not refer to anything. This proposal unifies the analysis of the autonomous verb in light of a lexically-restricted phenomenon, observed by McCloskey (2007), in which the autonomous verb is employed alongside an oblique experiencer to effect a construction similar to English raising via *it*-extraposition. Here, compare the non-autonomous construction in (3a) to its equivalent autonomous form in (3b) and the raising-like phenomenon in (3c).

- (3) (a) *Cheap mé go raibh cuma ghruama orthu.*
 think-PAST PRO-1SG that be-PAST look gloomy on them
 ‘I thought that they looked gloomy.’
- (b) *Ceapadh go raibh cuma ghruama orthu.*
 think-PAST-AUTO that they looked gloomy
 ‘It was thought that they looked gloomy.’
- (c) *Ceapadh dom go raibh cuma ghruama orthu.*
 think-PAST-AUTO to me that they looked gloomy
 ‘It seemed to me that they looked gloomy.’

Ultimately, what this analysis suggests is that the Irish language requires a syntactic subject to be expressed as either an external pronoun or as a verb ending, but that sentences without semantic subjects can be formed in Irish by using the autonomous ending on the verb as a pleonastic pronoun that fills the subject “slot” without contributing any additional meaning. Lexical-Functional Grammar is an ideal formalism for capturing this nuance because it independently expresses syntactic grammatical functions and semantic roles in a way that can be constrained by the lexical rules for each verb. In the above examples, a pronominal subject is given in (3a) while (3b) represents the productive usage of the autonomous form available to all Irish verbs. On the other hand, (3c) is the restricted phenomenon only available to verbs representing psychological states, where the syntactic subject is occupied by the autonomous verb form but there is an opening for a semantic experiencer, which in this case is taken by *dom*, ‘to me’. This three-way distinction between typical verb usage (3a), the productive autonomous form (3b), and the lexically-restricted raising-like phenomenon in (3c) will be a recurring focal point in this paper.

At the end of this first section, an overview is given for the computational approach applied in this paper. The Lexical-Functional Grammar framework (Falk, 2001; Kaplan & Bresnan, 1982) used in the present argument lends itself well to computational applications, and the present theoretical analysis is implemented exactly in a computer grammar that is capable of parsing even novel sentences within the distribution of its ruleset and lexicon. Section 2 of this paper examines previous analyses of the autonomous verb as something like an “impersonal passive” or as an analogous construction to indefinite pronouns in other languages. Later in this same section, the Lexical-Functional Grammar framework is put forward as being well-suited to frame this syntactic problem, and the argument of this paper is laid out: the subject of the autonomous verb in Irish is an expletive without a semantic role, and the various

constructions that use the autonomous verb are best handled as lexical rules. Section 3 presents a selection of data from the reviewed literature, discusses the implications of these attestations, and demonstrates the capabilities of a lexical-functional computer grammar written to parse the data and generate graphical output. Section 4 examines the implications of this paper's argument for analysis of Irish data and for the purposes of cross-linguistic comparisons.

1.1 LFG and Computational Applications

Lexical-Functional Grammar is a syntactic framework that analyzes sentences in terms of two structures: a constituent structure (C-structure), which expresses syntactic constituency hierarchically through tree diagrams, and a functional structure (F-structure), which uses attribute-value matrices to express grammatical functions, or relationships between elements of the sentence such as "subject" and "object". These two expressive notations (tree diagrams and attribute-value matrices) lend themselves well to computational applications, and computer grammars can be written that borrow techniques from unification grammar to parse novel sentences, so long as the contents of the sentences fall within the distribution of the grammar. A computer grammar along these lines was written to implement the theoretical conclusions of this paper using the Xerox Linguistic Environment (Crouch et al., 2008), a tool for LFG grammar engineering.

To illustrate the shape of an LFG computer grammar (echoing similar tutorial materials such as Falk (2001) and Bresnan (2000)), an example ruleset and lexicon are given in (4) that are capable of parsing the simple Irish sentence given in (2c). In a lexical-functional computer grammar, the ruleset defines the possible syntactic constituencies and the functional relationships they may express in the given language, while the lexicon lists the words covered by the grammar and their respective features.

When parsing, lexical rules are combined according to the rules for constituency - in the generated F-structure for a given sentence, each matrix or submatrix may contain one PRED value that indicates its meaning, some grammatical functions like SUBJ and OBJ that are specified within the PRED value, and any number of adjuncts (like prepositional phrases or adverbs) and features (like case, gender, or tense) that modify that level of the F-structure. In the lexical rule for the verb *buailim* (4c) below, the PRED value is given in single quotes to indicate a semantic value, the meaning is followed by a set of grammatical functions within angle brackets that are assigned semantic roles by the verb, and any grammatical functions that are specified syntactically but not assigned semantic roles may be placed outside the angle brackets (but within the single quotes).

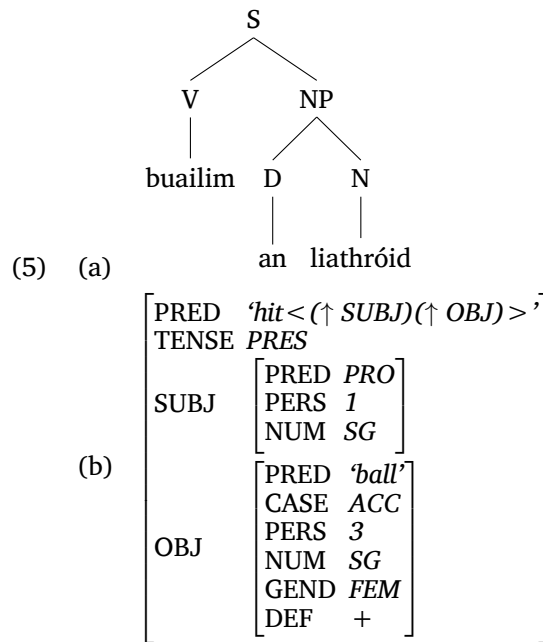
(4) Ruleset:

$$S \rightarrow \begin{array}{c} V \qquad \qquad \qquad NP \\ \left[\begin{array}{c} \uparrow = \downarrow \end{array} \right] \left[\begin{array}{c} (\uparrow OBJ) = ! \\ (\downarrow CASE) = ACC \end{array} \right] \end{array}$$

- (b) $NP \rightarrow \begin{matrix} DET & N \\ \uparrow=\downarrow & \uparrow=\downarrow \end{matrix}$
- Lexicon:
- (c) $buailim \rightarrow V: \left[\begin{array}{l} (\uparrow PRED) = 'hit < (\uparrow SUBJ)(\uparrow OBJ) >' \\ (\uparrow TENSE) = PRES \\ (\uparrow SUBJ \ PRED) = PRO \\ (\uparrow SUBJ \ PERS) = 1 \\ (\uparrow SUBJ \ NUM) = SG \end{array} \right]$
- (d) $an \rightarrow DET: \left[\begin{array}{l} (\uparrow DEF) = + \\ (\uparrow CASE) = ACC \\ (\uparrow GEND) = FEM \\ (\uparrow NUM) = SG \end{array} \right]$
- (e) $liathróid \rightarrow N: \left[\begin{array}{l} (\uparrow PRED) = 'ball' \\ (\uparrow CASE) = ACC \\ (\uparrow PERS) = 3 \\ (\uparrow NUM) = SG \\ (\uparrow GEND) = FEM \end{array} \right]$

In the ruleset given, C-structure rewrite rules such as (4a) should be familiar from other phrase structure grammar formalisms, but additional information is given below the child nodes to show their relationships in the F-structure. For example, the V node uses the rule $\uparrow=\downarrow$ to equate its information with that of the parent node S, while the NP node declares itself as the grammatical relation OBJ for the parent node. In the lexicon, the verb *buailim* (4c) is declared as a verb with the meaning ‘hit’ that assigns grammatical relations to a subject and an object. In this sentence, the verb also contains its own subject information, specifying the first-person singular pronoun as the subject. One other detail to pay attention to is that the case, number, and gender information for the definite article *an* (4d) and the noun *liathróid* (‘ball’) (4e) match, which is necessary for a successful unification of these two words into a noun phrase. In the ruleset, this unification is expressed in (4b) by the rule $\uparrow=\downarrow$ assigned to both the DET and N child nodes of the NP node, meaning that the features of both children must reside in the same submatrix of the F-structure.

In (5a) below, the resulting C-structure is given as a syntax tree while the accompanying F-structure is given as an attribute-value matrix in (5b). Again, note that the highest level of the matrix represents the information for both the S node and the V node that inherits it. In LFG terminology, this F-structure is both complete, meaning that all grammatical relations specified by the PRED value are present in the F-structure (here, SUBJ and OBJ), and coherent, meaning that no extra grammatical relations are present that are not specified by the PRED (which would be violated by, for example, assigning a direct object to an intransitive verb).



2 Analyses

2.1 Passive Analysis

One analysis that has been presented for the Irish autonomous verb is that it is a form of the passive. Nolan (Nolan, 2001) uses the term “impersonal passive” to refer to the autonomous, and treats it as a passive voice construction in which the agent of the verb is “specific but indefinite”. In other words, the information borne by the autonomous verb is that an agent exists, but it is not syntactically present. Nolan contrasts this with another construction in Irish, the “personal passive”, which is the cross-linguistically common passive construction seen in (2b) above. In this construction, a form of the verb *to be* is combined with a verbal adjective or verbal noun, and the object of the verb is promoted to the subject, expressed through the nominative case. This closely parallels the passive construction found in English and many other languages.

While this analysis is attractive in its simplicity, it raises a number of issues. First, although having multiple constructions to express the same thing is not impossible in a language, it seems improbable in this case that Irish would simultaneously have two very different but fully productive constructions that serve the same purpose of passivizing verbs. Nolan fails to make clear what the exact distinction in usage or meaning might be between the “personal” and “impersonal” constructions, since both imply the existence of an agent but don’t define it. Furthermore, the promotion of object to subject in autonomous constructions is not reflected in Irish case marking: the object

of an autonomous verb remains in the accusative case, which does not easily fit into the cross-linguistic definition of the passive. Again, it seems unlikely that Irish would have one construction that perfectly matches this idea of the passive and another that diverges in a significant way. In terms of morphology, the productivity of the autonomous form strongly contradicts the passive analysis, because intransitive verbs may also participate in the autonomous form, whereas a true passive construction requires the existence of a direct object to be promoted to subject position and/or nominative case. Finally, assigning a passive analysis to the Irish autonomous verb would simply not match the intuition of native speakers: the Christian Brothers, in their *Grammar of Irish* (Brothers, 1902), recount that Irish native speakers perceive the autonomous verb as being an active voice.

Nolan provides an important piece of data regarding the autonomous verb, however: examples of autonomous constructions have recently been attested in combination with an oblique agent. Typical passive constructions allow for an agent to be expressed obliquely because no semantic agent is otherwise expressed (e.g. *The ball was hit by me.*), while impersonal constructions do not because the impersonal pronoun has already taken this semantic role (e.g. the failure of **One hit the ball by me.*). Constructions like (6) are therefore significant because they show that while the subject's syntactic role is fulfilled by the autonomous verb, the semantic role of the agent is left unassigned and available for oblique expression.

- (6) *Caithfear* 1.39 milliún Euro á *caitheamh acu.*
 allocate-FUT-AUTO 1.39 million Euro for spending by them
 '1.39 million Euro will be allocated by them for spending.'

2.2 Analysis with *Arb*

McCloskey (2007) provides an extensive review of attested autonomous constructions in Irish and compares them to impersonal pronouns such as those used in German (*man*) and Italian (*si*). This would make the autonomous subject an impersonal pronoun and warrant parallel analysis to analogous constructions in other languages (including assignment of a semantic role). McCloskey's argument can be summed up with two core observations: first, that the autonomous subject may act as a controller for an open complement, and second, that it may be the antecedent of a certain reflexive pronoun. McCloskey acknowledges that the first of these observations is not enough to fully support his argument, noting that the English example *It was decided to go public* displays a control relation on the open complement *to go public* without a semantic subject as an antecedent. The second observation is therefore used to contrast the autonomous verb's purported ability to act as the antecedent to a reflexive pronoun with the failure of English **It was finally arranged for each other to be on the committee*. In other words, his argument hinges on support for the autonomous verb as the antecedent of reflexive pronouns.

McCloskey acknowledges that typical Irish reflexives, which use a pronoun plus *féin* (e.g. *é féin*, ‘himself’), cannot actually have an autonomous verb as an antecedent. His argument, therefore, relies on another, paraphrastic reciprocal pronoun *a/le chéile*, which McCloskey equates with English *each other*. This equation is not perfect, however. The phrase *a chéile* literally means ‘its/one’s companion’, and *le chéile* likewise means ‘with a companion’. In modern Irish, this phrase has been grammaticized into a pseudo-adverb meaning something close to ‘together’ or, in some cases, ‘each other’, but treating it strictly as a reciprocal pronoun is a restricted reading that does not necessarily match native speaker intuitions. In this light, McCloskey’s translation as given in (7) fails to capture that a non-reciprocal, non-grammaticized reading is perfectly available to native speakers.

- (7) *Théití ag ithe béile le chéile.*
 go-PAST-IMPERF-AUTO at eating a meal with a companion
 ‘One used to go eating a meal with a companion.’
 (or ‘People used to go eating a meal together.’)
 (cf. McCloskey, ‘People used to go for a meal with each other.’)

McCloskey also admits that examples like (7) may only occur under specific conditions, and that attested sentences are not universally accepted by native speakers when taken out of context.

The second issue with McCloskey’s argument arises when he introduces another construction that employs the autonomous verb and resembles *it*-extraposition for raising constructions in English, as seen above in (3c). The Government-Binding analysis requires a special, exceptional interpretation of (3c) vis-à-vis (3b) (i.e. *Ceapadh dom...* ‘It seemed to me...’ versus *Ceapadh...* ‘It was thought...’). Namely, McCloskey’s proposed feature *Arb* on the Tense node enters an agreement relation with a null pronoun in (3b), but in (3c) there must be a duplicate feature *Arb* already in place on the verb’s V node to accept the agreement relation from the Tense node, allowing an oblique experiencer to take the place of what normally would be a null pronoun. This analysis is unsatisfactory because it requires two separate analyses, the second of which simply assumes that an invisible feature is occasionally present on particular, lexically-determined verbs. This complex strategy attempts to account for a phenomenon that would be much more elegantly handled by a lexicalist approach, as detailed below in section 2.3.

One final consideration is that mentioned in section 2.1 above: an analysis which assigns a semantic role to the subject of the autonomous verb does not account for the availability of constructions with oblique agents, where it would attempt to assign the same semantic role twice. A successful and unified analysis of the Irish autonomous verb must therefore exclude the possibility of a semantic role being assigned to its subject. Note as well that the sentence in (6) is not part of the lexically-restricted phenomenon that McCloskey observes but a normal, fully productive instance of the autonomous verb.

2.3 Lexical-Functional Analysis

Lexical-Functional Grammar has been successfully applied to the Irish language before. Sulger (2009) demonstrated the effectiveness of the PREDLINK grammatical relation for connecting the related phenomena of copular constructions and topicalization in Irish. Namely, Sulger examined the effectiveness of different LFG formalisms for handling copular and substantive constructions in Irish (analogous to English sentences with *be*, *become*, etc.), and found that neither the single-tiered approach (which e.g. treats predicative adjectives as the heads of sentences rather than the substantive verb) nor the open-complement approach (which treats the subject of the sentence as a controller for e.g. predicative adjectives) best suits the Irish data. Rather, another grammatical functional called PREDLINK was proposed as a sort of closed complement function that can handle all of the available data on Irish topicalization, where even non-finite verb phrases may be fronted as a single unit out of finite clauses. This latter case is ungrammatical in English, where literal translation might yield examples like **It is painting a chair that the man was yesterday*. The same researcher later (Sulger, 2011) applied LFG again to analyze a notable subject-shifting phenomenon in Irish: the “be-at” possessive construction that is used in lieu of a verb meaning ‘to have’.

LFG lends itself well again to the present phenomenon. The functionalist approach means that verb forms such as the autonomous can be handled according to the grammatical relations they specify (translating to the constructions they may participate in), and the lexical approach means that phenomena like (3c) can be handled at the appropriate (lexical) level, since their productivity is determined on a case-by-case basis for each word. The autonomous verb in general, then, can be elegantly handled by Lexical-Functional Grammar as an expletive or impersonal subject, similar to that used in English weather verbs in phrases like *It rains*. Specifically, the lexical entries for verbs in the autonomous form should specify a subject grammatical relation that does not receive a semantic role (signified in LFG notation by placing it outside the angle brackets of the verb’s PRED value), and the subject’s information should be given in the lexical entry of the verb as ARB, indicating its impersonal status. Example lexical rules are given for an intransitive verb (8a) and a transitive verb (8b) in the autonomous form.

$$\begin{aligned}
 (8) \quad (a) \quad siúladh \rightarrow V: & \left[\begin{array}{l} (\uparrow PRED) = 'walk' <> (\uparrow SUBJ)' \\ (\uparrow TENSE) = PAST \\ (\uparrow SUBJ \quad PRED) = ARB \end{array} \right] \\
 (b) \quad buaileadh \rightarrow V: & \left[\begin{array}{l} (\uparrow PRED) = 'hit' < (\uparrow OBJ) > (\uparrow SUBJ)' \\ (\uparrow TENSE) = PAST \\ (\uparrow SUBJ \quad PRED) = ARB \end{array} \right]
 \end{aligned}$$

This approach avoids the several pitfalls faced by previous analyses. First, it matches native speaker intuitions: Brothers (1902) firmly state that Irish native speakers feel the autonomous is an active, not a passive, voice. Second, it elegantly handles the parallels between intransitive and transitive verbs in

the autonomous form rather than trying to divide their analysis like a passive interpretation would. It even fits into the broader scheme of Irish verbal morphology, as can be seen by comparing how the autonomous subject is specified in (8b) to (4c), where a different verbal suffix specifies a first person singular pronominal subject. Third, handling the autonomous subject within the lexical rule of the verb matches the restrictions that prevent it from participating in subject-shifting phenomena: these other constructions are those which cannot have an impersonal subject in Irish, such as possession and emotional states.

Most importantly, this approach allows for a strategy that avoids the asymmetry and complexity of the Government-Binding analysis with *Arb*, i.e. no silent or null nodes need to be proposed, no extra features need to be added to verbs in some cases but not others, and no additional rules need to be posited. Rather, a single additional lexical entry can be added to each of the handful of verbs that may participate in the raising-like phenomenon; these additions to the lexicon are warranted by the fact that, in many cases, the meaning of a verb can change somewhat depending on the construction it participates in (e.g. *ceapadh* translating to a form of ‘think’ in sentences like (3b), but to a form of ‘seem’ in sentences like (3c)). This flexibility is expressed in the parallel PRED values of the forms of the verb *ceap* ‘think’ given in the lexical rules (9a-9c) below.

- (9) (a) *cheap* → V: $\left[\begin{array}{l} (\uparrow \text{PRED}) = \text{'think'} < (\uparrow \text{SUBJ})(\uparrow \text{COMP}) > \\ (\uparrow \text{TENSE}) = \text{PAST} \end{array} \right]$
- (b) *ceapadh* → V: $\left[\begin{array}{l} (\uparrow \text{PRED}) = \text{'think'} < (\uparrow \text{COMP}) > (\uparrow \text{SUBJ})' \\ (\uparrow \text{TENSE}) = \text{PAST} \\ (\uparrow \text{SUBJPRED}) = \text{ARB} \end{array} \right]$
- (c) *ceapadh* → V: $\left[\begin{array}{l} (\uparrow \text{PRED}) = \text{'seem'} < (\uparrow \text{OBL})(\uparrow \text{COMP}) > (\uparrow \text{SUBJ})' \\ (\uparrow \text{TENSE}) = \text{PAST} \\ (\uparrow \text{SUBJPRED}) = \text{ARB} \end{array} \right]$

In application, these lexical rules mean that the verb form *cheap* will look for a syntactically external subject and closed complement, while the verb form *ceapadh* will trigger rule (9b) in cases where only a closed complement is present, while (9c) will be triggered when an oblique object is present as well as a closed complement. Analogous rules can be written for any of the verbs that participate in the raising-like phenomenon in modern Irish.

2.4 The Issue of Case

This paper proposes a functional approach that works well on both a theoretical level and in computational applications, but there is one remaining question that must be left for future work: Assuming no semantic role is assigned to the subject of the autonomous verb, why does the object of the verb remain in the accusative case? According to Dependent Case Theory (e.g., Marantz, 2000), the accusative case should only surface when a higher noun phrase is marked in the nominative case. An easy answer would be that Irish poses an exception

to the theory, but a deeper consideration would examine other possibilities. For example, native speakers may feel that because the autonomous verb is in the active voice, its object remains in the accusative to create contrast with the passive voice, which promotes the object to nominative case.

Another possibility is that the autonomous verb ending fills the same syntactic slot as a nominative pronominal subject, such that native speakers assign accusative case to the object based simply on its position in the sentence. This can be seen as parallel to native speakers of English who say *It's me*, with *me* in the accusative case based on its following the main verb, despite prescriptive grammars that argue that the correct case assignment would yield *It is I*. Regardless, the issue of case must remain for future work.

3 Data and Computational Analysis

This section examines a selection of attested modern Irish sentences and displays the results of parsing each with a computer grammar written for this paper in the Xerox Linguistic Environment. As introduced above, Lexical-Functional Grammar analyzes sentences according to a two-part framework, where constituency is expressed graphically via a syntax tree (the C-structure) while grammatical functions are expressed in an attribute-value matrix (the F-structure). The output of the computer grammar for each sentence is thus the combination of a tree and a matrix.

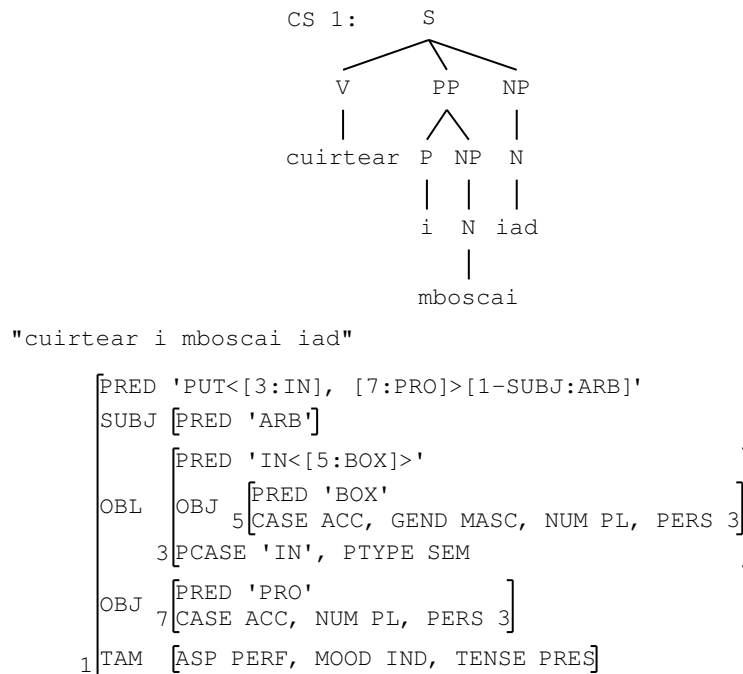
The first sentence to be examined is given in (10). This sentence demonstrates the basic usage of the autonomous with a transitive verb that is often translated into English using a passive construction. The pronominal object is clearly marked for the accusative case and postponed to the end of the sentence. This postponement is a syntactic option that modern Irish makes available only to direct objects, and it should be noted that *Cuirtear iad i mboscaí* would also be equally acceptable.

- (10) *Cuirtear i mboscaí iad.*
 put-PRES-AUTO in boxes PRO-3PL-ACC
 ‘(Someone) puts them in boxes.’

The output for sentence (10) is given in Figure 1 below. Note the “flat” C-structure for the sentence and the lack of a verb phrase node. Previous LFG analyses of Irish, which has V-S-O word order for finite clauses, have concluded that verb phrases are not needed for constituent analysis of the language (Carnie, 2005) (an acceptable position in LFG), although non-finite verb phrases do exist and can, for example, be topicalized. Thus, a typical Irish sentence will have a “flat” C-structure with the subject and object nodes’ being sisters to the main verb node. In the F-structure, the PRED value for the main verb assigns semantic roles to an oblique object and a direct object, and specifies but does not assign a role to the subject. Also note the sentence-level feature TAM, a set of the three features which Irish expresses on the main verb:

Tense (which may be past, present, or future), Aspect (which may be perfect or imperfect), and Mood (which may be indicative or conditional, with an external particle used to express interrogative status).

Figure 1: Output for sentence 10



The next sentence, given in (11), demonstrates an intransitive use of the autonomous. Examples like these would be difficult to translate into English if it were assumed that the autonomous verb is a passive construction, but assuming an impersonal subject allows an approximate translation with 'One.' This sentence also demonstrates one of the decisions that was made during the construction of the computer grammar. This was to abide by the modern spelling conventions for Irish (known as *An Caighdeán Oifigiúil* 'The Official Standard') as found in the *Foclóir Gaeilge-Béarla* (Ó Dónaill, 1977) for the purpose of laying the groundwork for a general-usage parser for the Irish language. The original spelling of the first two words, as given in McCloskey (2007), is *H-éirigheadh cleachtuighthe*, reflecting a dialectal spelling indicative of data collected in the field or from older sources. While a general purpose parser for Irish may one day be able to accommodate these alternative spellings, the present work had to prioritize using one spelling per word in order to reduce the amount of redundant work needed to parse each additional sentence.

- (11) *Éiríodh cleachtach le daoine a bheith ag*
 become-PAST-AUTO accustomed to people who be-INF at

teacht.
 arriving
 ‘(One) became accustomed to people arriving.’

Sentence (11) has its output given in Figure 2. In the C-structure, the shape of the tree is right-branching, demonstrating the head-initial typology that underlies a great deal of Irish syntax. This can also be found in the nested levels of the F-structure, where each is either specified by its parent level or adjunct to it. This analysis also used the PREDLINK grammatical relation suggested by Sulger (2009) for predicative adjectives in Irish (as discussed above in section 2.3).

Example (12) was discussed previously as an example cited by McCloskey for a reciprocal pronoun whose antecedent is the autonomous verb. Here, the translation ‘each other’ is used for *le chéile* with the caveat that it could just as easily be parsed as ‘with a companion’. The marking on the verb also demonstrates more of the range of the autonomous form. The verb is in the past imperfect, a meaning best expressed in English with ‘used to’. In total, there are five tense/aspect/mood combinations available in the autonomous: past, present, and future perfect, past imperfect, and the conditional.

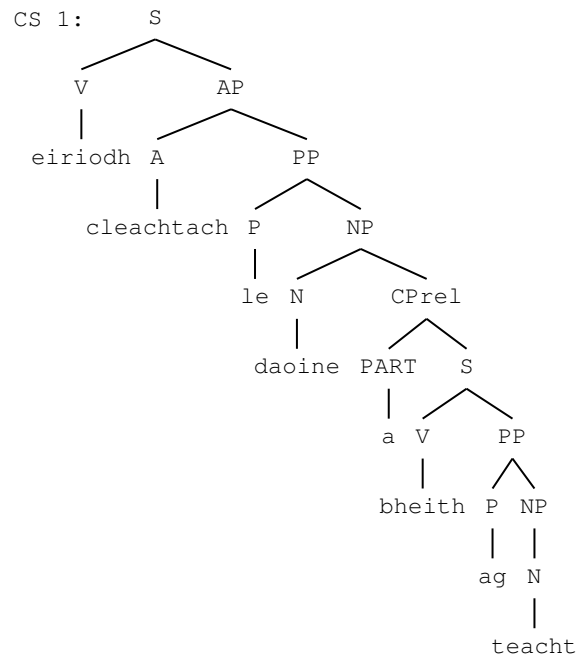
(12) *Théití ag ithe béile le chéile.*
 go-PAST-IMPERF-AUTO at eating meal with each other
 ‘(People) used to go eating a meal with each other.’

The output for sentence (12) is given in Figure 3, where an example of an Irish non-finite verb phrase can be found. Here, the sense is progressive because the verbal noun is within a prepositional phrase marked with *ag* and the direct object immediately follows the verbal noun. Elsewhere, in infinitive (i.e. non-progressive) verb phrases, the direct object immediately precedes a particle-marked verbal noun. A control relation can also be found in the F-structure, where the lexical entry for the verb specifies that its subject is the controller for the open complement it specifies (similar to English “(Someone) goes [verb]ing”). As noted in McCloskey, the autonomous subject may participate in control relations, but this is not an absolute indication of a semantic role because an indefinite subject can also be found as the controller for an open complement in languages such as English.

3.1 Parallel Constructions

This section is dedicated to presenting the ability of the computer grammar to precisely implement the lexical rules set out in (9a-9c) above for the parallel sentences found in (3a-3c). First, Figure 4 presents the parallel C-structures for the three sentences. Note the subject, which is expressed externally to the verb with a pronoun in the first tree, is expressed as an autonomous ending in the second and third trees. Likewise, note the comparison between trees two

Figure 2: Output for sentence 11



"eiriodh cleachtach le daoine a bheith ag teacht"

	PRED	'BECOME<[3:ACCUSTOMED_TO]>[1-SUBJ:ARB]'			
	SUBJ	[PRED 'ARB']			
		[PRED 'ACCUSTOMED_TO<[7:PERSON]>'			
		PRED	'PERSON'		
			{		
			PRED	'ARRIVE<[7:PERSON]>'	
			SUBJ	[7:PERSON]	
			TAM	[PROG +]	
			9	[PART-TYPE 'A', PCASE 'AT', PTYPE NONSEM]	
		7	[CASE ACC, GEND MASC, NUM PL, PERS 3]		
		3	[ATYPE PREDICATIVE, PCASE 'WITH-AGAINST', PTYPE NONSEM]		
1	TAM	[ASP PERF, MOOD IND, TENSE PAST]			

Figure 3: Output for sentence 12

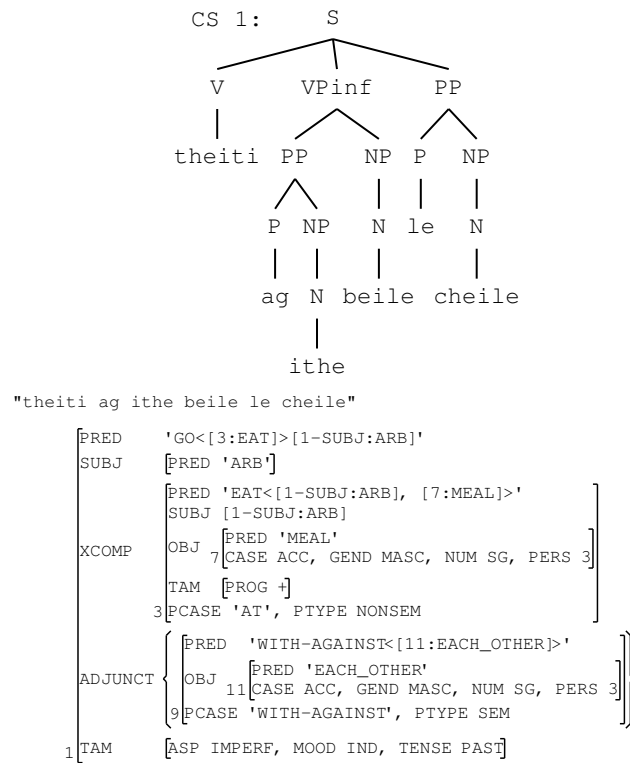
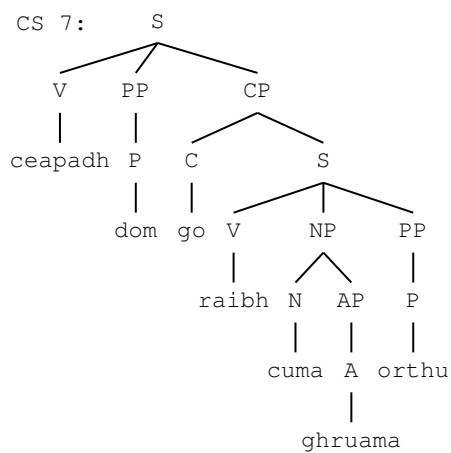
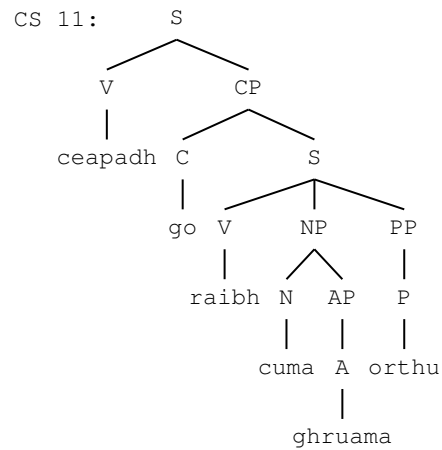
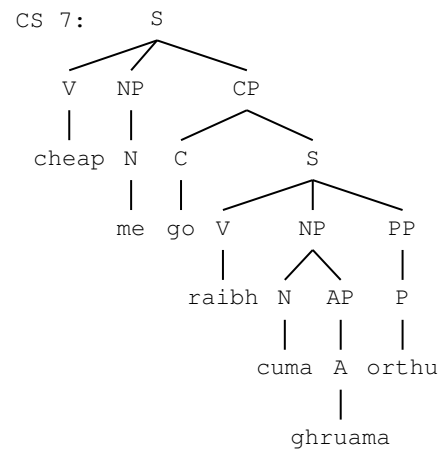


Figure 4: Paralell C-structures for sentences 10, 11, and 12



and three, which differ only by the presence of the oblique experiencer *dom* ‘to me’.

In Figure 5, the respective F-structures are given for the same sentences. Here, the parallelisms are even more clear: the verb in the sentence from (3 b) shifts its subject specification outside the angle brackets of its PRED value to indicate the lack of a semantic role. The sentence from (3 c) is treated similarly but also specifies an oblique experiencer as an additional grammatical function.

4 Implications of this Analysis

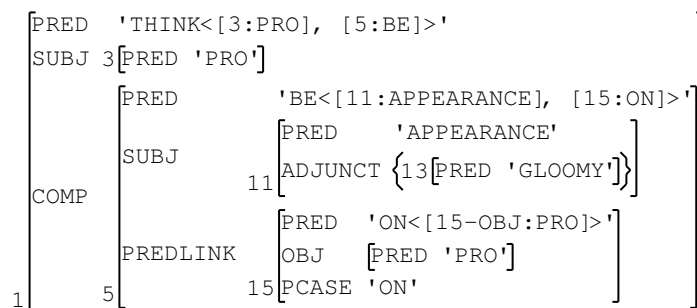
Expressed in LFG notation, the difference between a passive interpretation of the Irish autonomous verb and the interpretation proposed here may be significant, but the difference between an *Arb* analysis that assigns a semantic role to the subject of the autonomous verb and one that does not (the latter being the position of this paper) seems to simply be the difference between placing (\uparrow SUBJ) within or outside of the angle brackets for a given autonomous verb’s PRED value. The implications for analysis of Irish, however, are important. The analysis presented here is a unified account of the Irish autonomous verb across its entire productive range. What is revealed is a set of constructions that are uncommon among Indo-European languages but true to the expressive range of Irish: the impersonal subjects that languages such as English restrict to “weather verbs” and particular raising verbs are extended to include all verbs in the language, including both transitives and intransitives. While literary in comparison to modern linguistic formalisms, the Christian Brothers’ translation of “*Buailtear*.” as ‘A striking takes place.’ may be the best way to express in English this subjectless, objectless occurrence of a transitive verb that is perfectly acceptable to native Irish speakers.

The issue with the Irish autonomous verb in translation, however, is that it may not have one immediate equivalent in other languages. In English, for example, there is no construction that can convey an expletive subject with as much productivity as the Irish autonomous, which is productive for all verbs in the language. Capturing the meaning of the autonomous is also highly dependent on the verb in question: a passive-like translation may be used in English for the autonomous form of transitive autonomous Irish verbs, while an indefinite or non-specific subject is more appropriate for translating intransitive autonomous Irish verbs, e.g. ‘One walks.’ or ‘People walk.’ for “*Siúltar*.”

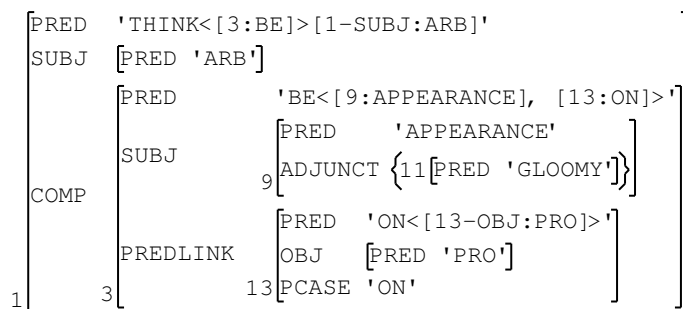
On the other hand, this very uncertainty plays into one of the strengths of LFG: the lexical component. If the varying “best translations” are treated at the lexical level, specific verbs can be coded for their meanings on a case-by-case basis. This is especially useful in the case of the same verb being used for different meanings based on its grammatical relations, as examined in section 3.1. There, the verb *ceap* ‘to think’ was treated as ‘It was thought’ when in the past autonomous form and accompanied by only a closed complement (‘that they looked gloomy’), but treated as ‘It seemed to (me)’ when in the same form and accompanied by both an oblique experiencer (‘to me’) and a closed com-

Figure 5: F-structures for sentences 10, 11, and 12

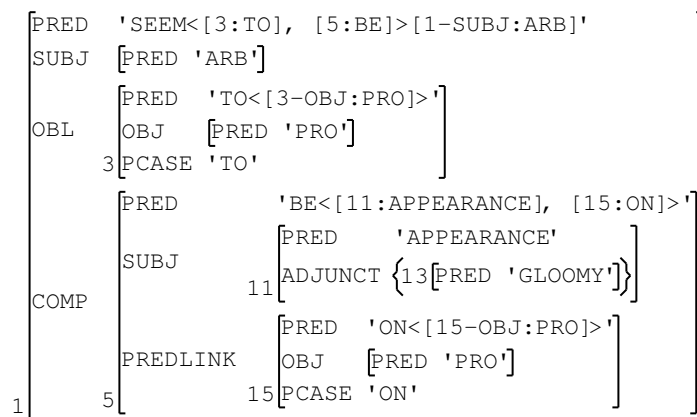
"cheap me go raibh cuma ghruama orthu"



"ceapadh go raibh cuma ghruama orthu"



"ceapadh dom go raibh cuma ghruama orthu"



plement. In other words, while a one-to-one correspondence to a phenomenon such as this may not be found in other languages, LFG provides the tools to handle the lexical variation that may occur in the meaning expressed by a given phenomenon when used with particular lexical items.

Finally, the computational component of the present research contributes to the wider field of natural language processing and its many applications. In particular, the construction of computer grammars for syntactic parsing enables applications such as predictive text assistants, chatbots, and machine translation. While recent work towards these applications has focused on statistical methods, the Irish language currently lacks the large, annotated training sets needed for deep statistical learning (see Judge et al. (2012) for a summary of the state of Irish with regard to digital resources and language technology). Hand-written computer grammars like the one constructed in the current paper can be applied in a limited fashion to these applications, but can also be used to annotate existing corpora and construct the kinds of datasets needed for statistical training.

5 Conclusion

The present analysis has been proposed to interpret the autonomous form of the verb in Irish within a lexical-functional framework and to account for related phenomena. Attested sentences were used as data to explore the range of constructions available using the autonomous verb, and this data was then used to engineer a computer grammar in the Xerox Linguistic Environment capable of parsing sample Irish sentences. A unified picture of the Irish autonomous verb as expressing an impersonal subject without a semantic role was formed, and the computer grammar effectively portrayed this in its automated analyses.

There is of course, more work to be done. The Irish language is generally under-studied in comparison to other languages in Europe, but it shows great potential for challenging and rewarding work in phonology, morphology, and syntax. In particular, Irish computational linguistics would greatly benefit from the construction of a morphological analyzer and tokenizer for the written language, and there are plenty of other syntactic phenomena in Irish to be analyzed. Potential topics of interest include relative clauses, which can be ambiguous in terms of subject versus object orientation and which may contain resumptive pronouns, and the syntax of copular sentences, which display complex word-ordering phenomena and mandate the usage of redundant pronouns in response to the presence of definite noun phrases.

References

- Bresnan, J. (2000). *Lexical-functional syntax*. Wiley-Blackwell.
Brothers, T. C. (1902). M.H. Gill.

- Carnie, A. (2005). Flat structure, phrasal variability and vso. *Journal of Celtic Linguistics*, 9, 13–31.
- Crouch, D., Dalrymple, M., Kaplan, R., King, T. H., Maxwell, J. T., & Newman, P. (2008). *Xle documentation*. Palo Alto Research Center.
- Falk, Y. N. (2001). *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. Center for the Study of Language; Information.
- Judge, J., Ní Chasaide, A., Ní Dhubhda, R., Scannell, K. P., & Uí Dhonnchadha, E. (2012). *The irish language in the digital space* (G. Rehm & H. Uszkoreit, Eds.). Springer.
- Kaplan, R. M., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). MIT Press.
- Marantz, A. (2000). Case and licensing. In E. Reuland (Ed.), *Arguments and case: Explaining burzio's generalization* (pp. 11–30). John Benjamins.
- McCloskey, J. (2007). The grammar of autonomy in irish. *Nat Lang Linguist Theory*, 25, 825–857. <https://doi.org/10.1007/s11049-007-9028-7>
- Nolan, B. (2001). Passive voice constructions in modern irish. *The ITB Journal*, 2, 51–78. <https://doi.org/10.21427/D76D1K>
- Ó Dónaill, N. (1977).
- Sulger, S. (2009). *Irish clefting and information-structure* (M. Butt & T. H. King, Eds.). Center for the Study of Language; Information.
- Sulger, S. (2011). *A parallel analysis of have-type copular constructions in two have-less indo-european languages* (M. Butt & T. H. King, Eds.). Center for the Study of Language; Information.