sciendo

SCP Volume 9 (2024)
Special Issue: Noun phrase and pronominal syntax in medieval and early modern Celtic languages
doi: https://doi.org/10.14746/scp.2024.9.7

ON THE SYNTAX OF THE LIGHT-HEADED NOUN PHRASES IN THE OLD IRISH GLOSSES: DEALING WITH MISSING DATA IN CORPUS LINGUISTICS [1,2]

CARLOS GARCÍA-CASTILLERO
https://orcid.org/0000-0002-4332-9872

University of the Basque Country (UPV/EHU)

ABSTRACT

The Old Irish glosses in contemporary manuscripts are the most reliable evidence for Old Irish syntax. These glosses convey discontinuous utterances that depend on the Latin text to which they

[1] This is a revised version of my contribution to the Workshop on Noun Phrase syntax of the Insular Celtic languages, held in Göttingen on the 26th-27th October 2023. This paper has profited from the comments of the participants at the Workshop and, in particular, from quite a number of relevant comments and corrections made by Elisa Roma and Elliott Lash. All errors and unaccuracies are my sole re-sponsibility.

[2] To the glosses established by the editors of this volume, I add the following ones: DECL = declarative clause type marker, LHD = light head, OBL.REL = oblique relative marker, REL = relative clause type marker, WH = interrogative pronoun. For the DECL gloss, I refer to García-Castillero (2020: Chapters 4 and 8). The Classes A,B of in-fixed pronouns are glossed as expressing declarative clause type, e.g. -sN- [3PL.DECL] in example (7), whereas Class C is glossed as expressing relative clause type, e.g. -id(L)- [3SN.REL] in example (1). For this interpretation, see García-Castillero (2020: 134-140). The elevated (L) in the previous form -id(L)- and others in the paper refers to the regularly expected mutation, which may not be expressed graphically or even linguistically.

Other abbreviations used in the main text are A (Agent of a transitive predi-cate), S (Subject of an intransitive predicate), P (Patient or object of a transitive predicated), Obl (Oblique constituent, in Old Irish, prepositional phrases), Gen (adnominal genitive case). For NP_{mat} and NP_{rel}, see Section 3.2.

For the analysis of the verbal complex, I deviate from the editor's indica-tions in cases in which a morphological boundary may reasonably be established.

are attached. One of the most obvious consequences of this discontinuous and textually dependent character is that the glosses very often convey what we could consider incomplete utterances, i.e. linguistic units that do not have the status of an independent clause or sentence.

When it comes to the study of NPs and their syntax, we face cases in which a gloss consists only of a bare NP. This paper deals with these isolated NPs, in particular, with those that are introduced by the light heads *intí*, *aní* and a^{N}, and defends the idea that these NPs constitute utterances that may be considered as complete, as against the initial impression, and therefore, that they may be considered for an inquiry on the syntax of Old Irish NPs.

This paper covers the following points: a basic quantitative description of the distribution of isolated light-headed NPs and the introduction of the notion of 'missing data' (cf. Osborne 2013: 128–129); a discussion of the notion of 'free NPs', i.e. apparently incomplete but meaningful linguistic units; an overview of the linguistic features of isolated NPs, in particular animacy, which is a statistically significant factor explaining the difference between isolated/dislocated light-headed NPs, which are more often animate, and intraclausal light-headed NPs, which are more often neuter; an outline of a wider investigation of all OIr. NPs in order to check if the above-mentioned statistical skew is in fact more general, i.e. that animate NPs are more likely in general to appear as free NPs.

Keywords: animacy, Old Irish, light-heads, incomplete NP, missing data.

1. Introduction: missing data in historical quantitative linguistics

The specific aim of this paper is the correlation between the lack of matrix clause of some NPs in the Old Irish glosses and a specific feature of these NPs. The NPs at stake are those introduced by a so-called light head, namely, *intí/aní* (eDIL s.v. 4 *í, hí*, CorPH s.v. *í* 1), and a^{N} (eDIL s.v. 4 *a*, CorPH s.v. *a* 4), and the feature considered is animacy.

The Old Irish glosses of the contemporaneous manuscripts often convey what apparently seems to be a syntactically incomplete utterance, so that we often have to deal with what is known as missing data in research fields such as social and health studies. It is in this context in which Osborne's (2013: 106) definition must be primarily understood: 'If any data on any variable from any participant is not present, the researcher is dealing with missing or incomplete data'. A general aim of this paper is to show that this notion of 'missing data' can be adapted to the corpus investigation of old languages, the quantitative historical linguistics of Jenset & McGillivray (2017), at least for this specific case of NP syntax.

This situation of missing data does not mean that Old Irish is a fragmentary language, in the sense established by Jürgen Untermann (1981). A fragmentary language suffers from a severe lack of data, and provides not enough of a basis for even the formulation of a complete phonological or morphological system, let alone a sufficient syntactic and lexical description. The Old Irish language of the glosses is not a fragmentary language in this sense, even though it has some structural *lacunae*, as we will observe later on.

In the elaboration of a dataset for an investigation on the NP syntax of the Old Irish language, which is in this paper limited to the light-headed NPs, we often

have to face the situation in which the gloss consists of a single NP with no matrix clause, i.e. with no verb on which this NP syntactically depends. This lack of matrix verb is the missing data in Osborne's definition above.

In the study of a language with a big corpus, the cases with missing data may well be simply ignored, one of the procedures noted in Osborne (2013: 117). For instance, Wiechmann's (2015: 65–70) study on Modern English relative clauses works with a database that requires, among other things, the complete syntactic context. Of course, a database of this type is just the desirable situation, and there is nothing wrong in such a procedure, but working with historical corpora is similar to social and health investigations in that the possibilities to get information may be fairly restricted. The more or less scarce sources of an old language are sometimes all that we can make use of, and – as a further circumstance which is essential to the traditional spirit of the studies on Antiquity – we are obliged to dismiss as few pieces of information as possible.

The paper is organized as follows. Section 2 introduces the corpus to be considered in this investigation, Section 3 defines the notion of light head and, thereby, the NPs that will be primarily investigated, and Section 4 describes the attestation types that are considered in the glosses for the NPs at stake. Section 5 offers a general quantitative description of these NPs, Section 6 focuses on the difference between neuter (i.e. inanimate) and masculine and feminine (i.e. animate) forms, and proposes various groupings to be considered in Section 7, in which the previous data sets are statistically tested. Section 8 discusses the results of the previous section, and Section 9 rounds off the paper.

2. The corpus

There are good reasons to restrict the corpus to the three great collections of glosses that are usually considered in studies of the Old *Irish* contemporaneous texts, namely, to the collections abbreviated as Wb., Ml. and Sg. that were edited by Strachan & Stokes (1901-1903) in their *Thesaurus Palaeohibernicum*. It is worth noting that partial or individual lexicographical elaborations of these three great texts are at the disposal of the reader, either in a traditional version, as Kavanagh (2001) for Wb., or in a digital version, as Griffith & Stifter (2013) for Ml., and Bauer & Schumacher (2015), Hofmann *et alii* (2023) for Sg. An important achievement in this sense is the *Corpus Palaeohibernicus* (= CorPH), i.e. Stifter *et al.* (2021), which includes Ml. and Sg. – but not Wb. –, as well as all other minor contemporaneous, and some other texts attested in later manuscripts.

For a corpus investigation on Old Irish syntax, and in this case in particular on NP syntax, the above-mentioned three big texts provide a number of attestations of quite a lot of linguistic elements that allow for drawing some reliable conclusions. The other, minor texts in contemporaneous manuscripts are not that extensive and

provide by far less attestations of light heads. Actually, for cases such as the one considered in this paper, even the Sg. collection of glosses constitutes a considerable problem mainly because it only provides two cases of the light head a^N, and makes marked use of the neuter forms of the light head *intí/aní*.

3. The Old Irish light heads *intí/aní* and a^N

Section 3.1 introduces the general notion of light head as a type of demonstrative form, and Section 3.2 the paradigmatic and functional range of forms under consideration in this paper.

3.1. The Old Irish light heads as demonstrative forms

The presentation of the Old Irish demonstratives in Table 1, taken from García-Castillero (2018: 167), aims at showing the structural place of the light head *intí/aní*, and includes neither the expression of distal demonstratives that involve the use of adverbial expressions such as *tall* 'there' or *ucut* 'yonder', for which I refer to Thurneysen (1946: 299-304), nor the use of stressed *so/siu* and *sin*, for which see Breatnach (2020). The demonstrative forms in Table 1 are based on the article, in this case, the nom.sg.m. *int-*.

Table 1. Old Irish adnominal and pronominal demonstratives (nom.sg.m. forms)

Syntactic use	Semantic use	Pragmatic use	ART	Noun / *-í-*	Particle / Proper name / Relative clause
Adnominal	Proximal	Situational / Exophoric	*int*	*salm*	*so* 'this psalm'
	Medial/distal	Anaphorical / Tracking	*int*	*salm*	*sin* 'that psalm'
	Neutral	Recognitial			(+ Proper noun)
		Determinative / Light head	*int*	*í*	(+ Relative clause)
Pronominal	Proximal	Situational / Exophoric	*int*	*í*	*siu* 'this one'
	Medial/distal	Anaphorical / Tracking	*int*	*í*	*sin* 'that one'

The system described in Table 1 allows for a general definition not only of the Old Irish light head *intí/aní*, but also of the light head a^N: it is a pronominal demonstrative with neutral deictic meaning which is followed by a relative clause. Diessel (1999: 135-139) uses the term 'determinative' for this type of demonstrative which 'functions to mark the nominal head of a relative clause'.

The term 'light head' is based on Citko's (2004) and Epps' (2009) contributions. Citko (2004) applies the term 'light head' to Polish demonstrative forms serving to introduce relative clauses.[3] This term has been applied by Lash (2011: 215-216) to the Old Irish form *a*^{N}. Old Irish differs from the languages considered by Citko and Epps in that it rarely admits headless relatives, so that the forms which are termed 'light heads' represent the default means of making an NP out of a relative clause.[4]

For other pronominal uses in Table 1 in which the form *intí/aní* has a specific spatial meaning in combination with the particles *-siu* (proximal) and *-sin* (distal), see Thurneysen (1946: 301). For the 'recognitional' use of *intí/aní* with a name, e.g. *inti moysi* 'the well-known, the aforementioned, this Moses' (Wb. 33a2), I refer to García-Castillero (2018: 165-167).

3.2. Paradigmatic constituency and functions of the Old Irish light heads

The pretonic light head *a*^{N} is a paradigmatically isolated form expressing only nom./acc.sg.n. case without preposition,[5] whereas the stressed *intí/aní* has a full-fledged paradigm given in Table 2, which offers the most frequent variants, and gives the feminine forms that are at least once attested.[6]

---

3 For more instances, in this case from the Teramano dialect of Italian, see Mantenuto & Caponigro (2021).

4 Ó Cathasaigh (1990) offers a number of relative clauses without antecedent in both the texts of the corpus and, more often, in other texts attested in non-contemporaneous manuscripts. In a study on precisely this type of relative clause in Old Irish, Roma (2023) offers 17 examples from the CorPH as assured cases of headless relative clauses, and 7 further examples as doubtful cases. As also noted by Roma, some of the examples given as assured by her have been interpreted in a different manner by Stokes & Strachan (1901-03), precisely assuming a missing or omitted light head. The existence of headless relative clauses in Old Irish cannot be denied, though it is also necessary to acknowledge that they are really scarce, at least in the contemporaneous texts. For reasons of space, I refrain to discuss this point.

5 As Elisa Roma kindly points out to me, one of the NPs introduced by *a*^{N} in the corpus is used in a context (i.e. after a verbal noun) in which genitive case would be expected: […] et *nebchretem anadiadar di-*crist (neb-cretem a^{N}=ad·(f)iad-ar di-crist) [not-believing.NOM LHD.ACC.SG.N= PV·(REL)declare.PRES.IND-3SG.PASS of-Christ.DAT] '[…] and not believe what is declared of Christ' (Wb. 27a10). In my view, this represents an exception to the rule that requires genitive case for the object appearing after a verbal noun, and is probably based on structures mentioned in footnote 9 below, in which the light-headed NP precedes the verbal noun. This form is counted as accusative.

6 Most probably, in view of the form of the article, the unattested but assumable feminine forms would be as follows: gen.sg.f. *innahí*; nom./acc.pl.f. *innahí*, gen.pl.f. *innaní*, and dat.pl.f. *-(i)naibhí*.

Table 2. The paradigm of the Old Irish light head *intí/aní*

	Singular			Plural		
	Masc.	Fem.	Neut.	Masc.	Fem.	Neut.
Nominative	*intí*	*indí*	*aní*	*indí*	?	*innahí*
Accusative	*inní*, *-(s/i)nní*	*inní*	*aní*, *-(s)aní*	*innahí*, *-(s)nahí*	?	*innahí*, *-(s)nahí*
Genitive	*indí*	?	*indí*	*innaní*	?	*innaní*
Dative	*-(si)ndí*	*-(si)ndí*	*-(si)ndí*	*-(s)naibhí*	?	*-(s)naibhí*

The forms *aní* and a^{N} coincide in their character of nom.sg.n. and acc.sg.n. without preposition; the different use of each form is analysed in García-Castillero (in prep., Chapt.2). For every other case (included the accusative with preposition) of the Old Irish nominal declension, only the forms of the paradigm of *intí/aní* are used. For the scarcity of feminine forms, see Section 6.1 below.

The use of the light heads as antecedents of relative clauses has the consequence that they potentially have both a function within the matrix clause (i.e. the NP_{mat} function), and a function within the relative clause that comes after (i.e. the NP_{rel} function). The notation $NP_{rel/mat}$ may be found in Andrews (2007: 206).

The light head *intí/aní* in Table 2 expresses all $NP_{mat/rel}$ functions of the Old Irish nominal inflection, which are A, S, (= nom.), P (= acc.), Obl (= preposition + acc./dat.), and Gen. (= gen.). By contrast, the light head a^{N} only expresses A, S, P functions in the corpus.[7]

Example (1) includes two forms of the nom.sg.f. form *indí*, which refer to the feminine noun *persan* '(grammatical) person' previously mentioned in the gloss.

(1) […] .i. *is sain indí asidrubart 7 indí frisanérbrath* (Sg. 220a10)
.*i.* is·sain indí
i.e. COP.PRES.IND.DECL.3SG·different/NOM.SG.F. LHD.NOM.SG.F
as-id·rubart 7 indí
PV-3SN.REL·PRF.say.PRET.ACT.3SG and LHD.NOM.SG.F
fris-a^{N}·érbrath
to-OBL.REL-PRF.say.PRET.3SG.PASS
'different are the one who has said it and the one to whom it has been said'.

In example (1), the first *indí* has S NP_{mat} – A NP_{rel} functions, whereas the second *indí* has S NP_{mat} – Obl NP_{rel} functions. The functional profile of other cases quoted below is the following: in (7), A NP_{mat} – A NP_{rel}; in (8), S NP_{mat} – A NP_{rel}; in (9), Gen NP_{mat} – A NP_{rel}; in (10), Obl NP_{mat} – S NP_{rel}.

[7] See Thurneysen (1946: 298) for the rare use of this form a^{N} after a preposition in texts other than those of the corpus.

4. Attestation types of the light-headed NPs in the glosses

The corpus includes 635 tokens of the light heads *intí/aní* and *a*^{N} described in the previous section, distributed as showed in Table 3. For a more detailed numerical distribution of attestation types, see Tables 4 and 5 in Section 5 below.

Table 3. Forms of *intí/aní* and *a*^{N} used as light head in the corpus

	Wb.	Ml.	Sg.	Total
intí/aní	111	258	138	507
a^{N}	98	28	2	128
Total	209	286	140	635

The token counts in Table 3 include every form found in the corpus, but not all of them are accompanied in their text by the minimal context that would be expected in an assumably complete utterance. This is why a classification of four main relevant attestation types is proposed: (a) isolated form, (b) isolated NP, (c) dislocated NP, (d) intraclausal NP. These types are discussed in turn in this section.

4.1. Isolated forms of light heads

There are twenty-four cases of isolated forms in the corpus, all of them belonging to the light head *intí/aní*. The light head *a*^{N} is not found in this attestation type.

Most often, the light head appears as the sole form in the gloss, typically attached to a relative form of the Latin text. For instance, *indí* [nom. pl. masc.] (Wb. 1c6), is used to gloss Latin *qui*. Since Old Irish does not express relative subordination by means of an independent relative marker or pronoun, as stated in Section 3.2 above, it seems that the glossators often felt the need to make clear that the Latin expression involved an NP for which in Old Irish not only the relative verb, but also the light head was normally used. As a further possibility, the form *aní* [nom./acc. sg. neut.] (Sg. 182a3) glosses a Latin noun in order to state that the word is being quoted; see Section 4.3 below for this issue.

In some few cases of isolated forms, the light head *intí/aní* is preceded by other elements, but it is not followed by a relative verb.[8]

---

[8] I find three such cases in the corpus. One is the first form in *amal bid innani* .i. *innaní dufurgabat* (amal bi-d·innani .*i.* innaní du·$^{(L)}$furgab-at [as COP.PAST.SUBJ-3SG.DECL·LHD.GEN.PL.M i.e. LHD.GEN.PL.M PV·(REL)raise.PRES.IND-3PL.ACT]) 'as if it were of those, i.e. of those who raise' (Ml. 75a2-3), in which the second form is interpreted as a case of isolated NP. The second is *cionaibhi* (cio-naibhi [WH-LHD.DAT.PL]) 'by which?' (Ml. 47a10), translating Latin interrogative *quibus*, literally 'what is it by those […]?'. The third is *nadndenat indidail inna hí* (nad·^{N}den-at ind-idail innahí [NEG.REL·^{REL}do.PRES.IND-3PL.ACT ART.NOM.PL.M-idol.NOM.PL LHD.ACC.PL.N] 'that the idols do not do those (things)' (Ml. 37b26), which is edited in this

The two glosses Sg. 5a6 and Ml. 108b4, in which the light head is directly followed by a Latin relative pronoun (*qui reliqua* and *quos solet* respectively), have been classified as isolated forms, and not as intraclausal light-headed NPs.

4.2. Isolated light-headed NPs

An isolated light-headed NP implies that the NP either appears alone in the gloss or, if it appears with other elements in the same gloss, that it cannot be considered as dislocated or intraclausal. This type has as much as 216 cases in the corpus.

Quite a lot of these cases are more or less literal translations of a Latin expression, as (2), which offers the whole gloss referred to Latin *qui non manducat* 'the one who does not eat'.

(2) .i. *intí óinas* (Wb. 6b19)
.*i.* intí óin-as
i.e. LHD.NOM.SG.M fast.PRES.IND-3SG.ACT.REL
'the one that fasts'.

(3) .i. *indí crettes cen imdibe storide* (Wb. 2b19)
.*i.* indí crett-es
i.e. LHD.GEN.SG.M believe.PRES.IND-3SG.ACT.REL
cen imdibe$^{(N)}$ storide
without circumcision.ACC material.ACC.SG.N
'of him who believes without material circumcision'.

Some of these isolated NPs, however, are not bare translations of Latin relative clauses, e.g. the gloss in (2), which is attached to *per fidem* 'through (the) faith' and states the person who possesses that faith.

The ascription to this type is not always self-evident, and I have to insist in that the basic criterion is that the NP at stake is neither intraclausal nor dislocated. This is why some light-headed NPs are taken as isolated, e.g. the case in (10), an example discussed in Section 8.1 below, though they may still be considered in a specific linguistic context.

---

manner by Griffith & Stifter (2013) and in the CorPH, but not by Stokes & Strachan, who read *inna hí nadndenat indidail*. Griffith & Stifter's (2013) arguments, i.e. the specific palaeographic situation (*inna hí* is written under *nadndenat indidail*), and the tendency of relative nasalization in restrictive relatives to appear only after singular animate antecedents (which is complemented by the still clearer tendency of the light head *intí/aní* to be followed by relative lenition), seem to me convincing, even though this attestation type (namely, the light head with previous context but without relative verb) is not frequent at all. One could perhaps think that either *-sin* or *-siu* (see Table 1 above) has been omitted after the form *inna hí*; in this case, it is then a pronominal demonstrative 'of those, of these (things)', and not a light head.

4.3. Dislocated light-headed NPs

A (left- or right-)dislocated NP is a syntactically detached NP which is somehow referred to or implied by another element of a clausal unit with which the detached NP constitutes a sentential unit. Table 4 below shows that there are more cases of left-dislocated light-headed NPs (n= 64) than of right-dislocated ones (n= 18) in the corpus.

Left-dislocated NPs appear only in nominative case (the *nominatiuus pendens* of the traditional grammar), as in (4), which is also an example of the use of the neuter forms of *intí/aní* in quoting a Latin word. For this and other manners of quoting words in Old Irish, see Breatnach (1990). Right-dislocated NPs can also be marked by other cases of the Old Irish nominal inflection, as in the gloss of (5), in which the NP introduced by the gen.pl.m. light head *innaní* is attached to an NP also in gen.pl.m. form, and explains it.[9]

(4) […] *Aní as* cognomen *leosom is* nomen *són leissem* (Sg. 28a18)
aní as$^{(L)}$·*cognomen*
LHD.NOM.SG.N COP.PRES.IND.REL.3SG·*cognomen*
le-o=som is·*nomen*
with-3PL=EMPH(3PL) COP.PRES.IND.DECL.3SG·*nomen*
són leis=sem
ANA with.3SM=EMPH(3SM)
'that which is *cognomen* for them is *nomen* for him [i.e. for Priscian]'.

(5) .i. *nu inna turcabthae* .i. *innaní soirtar and* (Ml. 75a4)
.i. nu inna$^{(N)}$=turcab-th-ae *.i.*
i.e. or ART.GEN.PL.M=raise-PPL-GEN.PL.M i.e.
innaní soir-tar and
LHD.GEN.PL.M deliver.PRES.IND-PASS.3PL.REL there
'or of the raised, i.e. of those who are delivered there'.

(6) […] .i. *matuhé ata horpamin indí rochomalnisset recht* (Wb. 2c14)
.i. ma-tu·hé ata$^{(L)}$·horpam-in
i.e. if-COP.PRES.IND.3PL·3PL COP.PRES.IND.REL.3PL·heir-NOM.PL
indí ro·^{L}/comalni-ss-et recht
LHD.NOM.PL.M PRF·REL/fulfill-PRET-3PL.ACT law.ACC
'if they are heirs who have fulfilled the Law'.

---

9 As cases of left-dislocated NP introduced by a light head I count three cases (Wb. 5c23 with *a*^{N}, and Ml. 29a8, Ml. 31b24 with *aní*) in which the NP at stake appears before a verbal noun: e.g. […] et *aforchongair dochomalnad* (*et* a$^{(N)}$=for·^{L}congair do-chomalnad [and LHD.NOM.SG.N= PV·^{REL}/order.PRES.IND.3SG.ACT to-fulfilling.DAT]) '[…] to fulfil what he orders' (Wb. 5c23).

Example (6) illustrates a case in which the NP introduced by the nom.pl.m. *indí* is interpreted as a right-dislocated NP co-referrential with the 3pl pronoun *hé* of the conditional clause, which is properly a cleft-sentence (i.e. 'if it's they who are heirs, the ones who […]').

This is the place to note, in order to anticipate the discussion in Section 8.3 below, that Old Irish makes no distinction between the two structures of fronting termed in the literature as topicalization and left-dislocation. In a language such as English, and using the examples given by Foley (2007: 443), a case such as *That dish, I haven't tried* is considered topicalization, due to the lack of pronominal resumption in the main clause, whereas *Thai cooking, I find it irresistible* is a case of left-dislocation. As stated in García-Castillero (2020: 78), Old Irish does not distinguish these two constructions, there being basically one type, which aligns rather with the left-dislocated pattern of the previous distinction. To be more precise, the case in which the left-dislocated NP is the subject of the main verb in the tautosentential clause has the same form as if the NP would be in an intraclausal position, as e.g. in (4) above. For the case in which the left-dislocated NP is the object of the tautosentential main clause, the verbal complex in the latter bears the corresponding object pronominal marker; see García-Castillero (2020: 79-80) for examples.

4.4. Intraclausal light-headed NPs

Intraclausal NPs in Old Irish follow their main verb as a part of the argument structure of the clause, there being no assumable pause or syntactic disruption before them, in which case we should understand a case of right dislocation, as in (5) and (6) above. I count a total of 314 cases of intraclausal light-headed NPs in the corpus. Apart from the case in (1) above, this intraclausal type is illustrated in the two following cases.

(7) .i. *dosnucci intí dodarogart* (Wb. 22c1)
.i. do-s^{N}·ucci intí
i.e. PV-3PL.DECL·understand.PRES.IND.3SG LHD.NOM.SG.M
do-da·rogart
PV-3PL.REL·PRF.call.PRET.ACT.3SG
'He who has called them understands them.'

(8) .i. *bithé magistir dongegat indhí asindisset atola féisne dóib* (Wb. 30d8)
.i. b-it·hé magistir
i.e. COP.FUT-3PL.DECL·3PL master.NOM.PL
do·^{N}-geg-at indhí
PV·REL-choose.FUT-3PL.ACT LHD.NOM.PL.M

as·(L)indiss-et a=tol-a
PV·(REL)declare.FUT-3PL.ACT 3PL.POSS=desire-ACC.PL
féisne dó-ib
REFL to-3PL
'the masters whom they will choose will be those who will declare to them their own desires'.

In example (7), the NP introduced by the nom.sg.m. *intí* is the agent of the preceding verb, and both main and relative verb express the same pronominal object reference, though with different affixal forms, the Class A (for declarative) and Class C (for relative clause types) forms of infixed pronouns respectively.

In (8), the copular clause conveys a referential predicate in which the referent of the Latin *magistros* is further specified with an NP. Though this structure is quite similar to the one in (6) above, to which it is most probably diachronically related, as suggested in García-Castillero (2020: 229-332), I assume that (8) counts as a case of non-verbal referential predication in which the NP *magistir dongegat* 'the masters whom they will choose' is equated to *indhí asindisset…* 'those who will declare …'. In Old Irish, this type of non-verbal predication is introduced by the grammaticalized use of the copula and the tonic pronoun (i.e. *bithé*). The number of these referential non-verbal predicates with *intí/aní* is very reduced in the corpus: I find just another one (i.e. Ml. 54a12), from the total of 314 cases of intraclausal light-headed NPs.[10]

4.5. Final notes on the attestation types of the Old Irish light-headed NPs

Some light-headed NPs in the corpus could certainly be included in more than one of the proposed four types of the previous classification, and this is probably due to the similarities that these types may have and, in the specific case of the Old Irish glosses, to the linguistically hybrid character of the texts as well.

The Old Irish scholars made relatively free and spontaneous use of the Latin language in combination with their own language in not a few cases, and they surely considered a hybrid clause containing both Irish and Latin as one single message. See Bisagni (2013-14), with previous references, for this type of code-switching in the Old Irish glosses.

In a more linguistic vein, one should not assume that NP isolation, at least in the negative definition put forward in this classification, equals complete lack of any linguistic context. Moreover, as discussed in Section 8.1 below, some of the isolated NPs of the corpus resemble dislocated NPs. Finally, the discussion on the difference between examples (6) and (8) above also makes clear that the

[10] As cases of right-dislocation, that is to say, as parallel to (6), are interpreted Wb. 4c40, Ml. 56a20, Ml. 46c14, Ml. 111c13, Ml. 114c7.

difference between right-dislocated and intraclausal NPs is not that clear as one may assume at first sight, at least in Old Irish.[11]

Section 8.2 below will elaborate on the relationship between these attestation types, in particular on the links between the isolated and dislocated NPs, which constitute the main topic of this paper.

5. General quantitative description

Table 4 below offers the light heads *intí/aní* and a^N presented in Table 3, divided by case, number and gender, and according to the attestation types established in the previous section.

Given that there is no formal distinction between nominative and accusative (singular and plural) in neuter elements, the case assignment for the right-dislocated and intraclausal NPs of these forms is made on the basis of their S/A or P NP_{mat} function, if the form has no preposition. If the neuter form *aní* is introduced by a preposition, then it is of course taken as accusative.

Table 4. Gender, case and number in the light heads *intí/aní* and a^N according to attestation types

		Wb.			Ml.			Sg.		
		intí	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N
		m./f.	n.		m./f.	n.		m./f.	n.	
Isolated forms	Nom.sg.				2	4			1	
	Acc.sg.				1					
	Gen.sg.				2					
	Dat.sg.				1					
	Nom.pl.	1			1	1		1		
	Acc.pl.				3	3				
	Gen.pl.				1				1	
	Dat.pl.					1				
Total of isolated forms		1	0	0	11	9	0	1	2	0
Isolated NPs	Nom.sg.	10	4	13	9	5	7	3	19	
	Acc.sg.		2		7	1				
	Gen.sg.	1			11	3			13	
	Dat.sg.	2			4	3			29	
	Nom.pl.	8	3		16	5			5	

[11] A syntactic feature of the NPs introduced by the light heads *intí/ani* and a^N is that they are not used as the focused constituent of the cleft-sentence.

	Acc.pl.	2			5					
	Gen.pl.	1			7					
	Dat.pl.				15			1	2	
Total of isolated NPs		24	9	13	74	17	7	4	68	0
Left-dislocated NPs	Nom.sg.	13	8	12	9	9	1		2	
	Nom.pl.	6			1	2			1	
Total of left-disloc. NPs		19	8	12	10	11	1	0	3	0
Right-dislo-cated NPs	Nom.sg.			3	4	1				
	Acc.sg.			3		1				
	Gen.sg.									
	Dat.sg.									
	Nom.pl.	3			1					
	Acc.pl.	1								
	Gen.pl.				1					
	Dat.pl.									
Total of right-dislocated NPs		4	0	6	6	2	0	0	0	0
Intra-clausal NPs	Nom.sg.	7	14	32	11	22	14	4	7	1
	Acc.sg.	1	4	35	5	38	6	1	3	1
	Gen.sg.	1	1		1	3			12	
	Dat.sg.	1	2		1	15		3	26	
	Nom.pl.	2			4	4				
	Acc.pl.	4	1		4	4			2	
	Gen.pl.	6			2				1	
	Dat.pl.	2			3	1			1	
Total of intraclausal NPs		24	22	67	31	87	20	8	52	2
Total 1		72	39	98	132	126	28	13	125	2
Animate vs inanimate		72	137		132	154		13	127	
Total of *intí/aní* vs a^N		111		98	258		28	138		2

For the isolated NP type, the distinction between nominative and accusative cases for neuter forms depends on the combination with a preposition: if there is no preposition, then the form is intended as nominative; if there is preposition,

then it is understood as accusative. Recall that the combination with a preposition is not possible for the light head a^N, which is only taken as a nom./acc.sg.n. form.

At the bottom of Table 4, the masculine and feminine forms of *intí/aní* are united in one single column, which is labeled as 'animate'. See Section 6.1 below for a justification.

The distribution of the forms in Table 4 allows for a number of observations, among others, the one that constitutes the main point of this paper. In this section, however, I focus only on the most general ones, that may be gleaned from the bottom of the table.

First, the frequency of the light head a^N is lower towards the right-most part of the table. Since Ml. and Sg. are attested a century later (in 850/1 AD) than Wb. (750 AD), one is tempted to assume a diachronic decrease in the use of a^N. This may well be true, but only if Wb. is compared to both Ml. and Sg. taken as a whole. Actually, it is in Sg. where the use of a^N is rare, most probably due to the proper character of the text, in which Latin words are often quoted. The rule seems to be that only the neuter forms of *intí/aní* are used for this quoting use. Example (4) above is such a case, with left-dislocated NP, but it has also been noted in Section 4.1 that some isolated forms can be understood in this manner, and the two other attestation types in Sg. are mostly used with this quoting function: this is the case of the 68 cases of isolated NPs and of the 52 intraclausal NPs in Sg. Though I think that this quoting use is also linguistically significant, it is clear that the situation of the light-headed NPs in Sg., with its prevalence of neuter forms, is to be attributed to the specific quoting use,[12] which is expected in commentaries to a grammatical text such as Priscian's *Institutiones Grammaticae*. This is why Sg. must be taken out from the analysis of this paper that looks at the different use of the animate (i.e. masculine and feminine) light heads with respect to the inanimate (i.e. neuter) ones.

Second, and in spite of the previous general observation on a^N, neuter forms (counting both those of *intí/aní* and a^N) are more frequent than masculine and feminine forms in the corpus. Once again, the extreme preponderance of neuter forms in Sg. is surely to be attributed to the above quoting use.

The lack of a^N in the so-called 'isolated forms', already mentioned in Section 4.1 above, is surely due to the fact that this light head a^N is a pretonic element that cannot appear by itself, whereas the forms of *intí/aní* are stressed and, therefore, capable to be uttered by themselves. Though graphical independency must not be inevitably based on prosodic independency, this is still a case in

[12] From the total of 135 non-isolated Sg forms of *intí/aní* that are given in Table 5 below, no less than 102 are used in this quotation use that is exemplified in (4) above. That is to say, be it an isolated, dislocated or intraclausal NP, the NP has a neuter light head followed by the relative form of the copula plus a Latin word or, on less occasions, a Latin expression.

which an undoubtedly clear case of defective attestation obeys to a linguistic reason of some type.

Be that as it may, one of the steps in the process of filtering the raw quantities of Table 4 is the elimination of the cases of 'isolated forms' from the dataset to be used in the next sections. These figures are given in Table 5, which includes the total quantities of the three collections of glosses.

Table 5. The distribution of gender in the light heads *intí/aní* and a^N (excluding the isolated forms of Table 4)

	Wb.			Ml.			Sg.			Total		
	intí	*aní*	a^N	*in*	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N
	m./f.	n.		m./f.	n.		m./f.	n.		m./f.	n.	
Total 1	71	39	98	121	117	28	12	123	2	204	279	128
Anim. vs. inanim.	71	137		121	145		12	125		204	407	
intí/aní vs a^N	110		98	238		28	135		2	483		128

Whereas there is a total number of 635 forms belonging to the light heads *intí/aní* and a^N in Tables 3 and 4 above, Table 5 involves a total of 611 forms, that is to say, the 24 cases of 'isolated form' of light heads found in the corpus are not further considered in this study.

6. Gender distribution according to the attestation type in the corpus

6.1. For a classification in terms of animate vs inanimate light heads

Bearing in mind only the cases in Table 5 above, the set of attested feminine forms is reduced to one example in Wb. (out of 208 cases = 110 of *intí/aní* + 98 of a^N), two in Ml. (out of 266 = 238 + 28), and two in Sg. (out of 137 = 135 + 2). These five feminine forms represent the 0.82% of the total amount of light heads in Table 5, i.e. 5 cases out of 611. Given that the light head *intí/aní* is based on the form of the article, the assumable formal distinction between the feminine and the other genders is only to be expected in the nom.sg. (fem. *indí* vs masc. *intí*), gen.sg. (masc./neut. *indí* vs. expected fem. *innahí*) and in the nom.pl. (fem./neut. *innahí* vs. masc. *indí*).[13] The only form of them which is attested in the glosses is

[13] This seems to be a case of polarity in the marking of gender, in line with Corbett's (1991: 196) explanation of the Somali article, which is *-kii* for masc.sg. and fem.pl., and *-tii* for masc.pl. and fem.sg. In the present case, *indí* is the light head for gen.sg. masc./neut. and nom.pl.masc., whereas *innahí* is the light head for gen.sg. fem. and nom.pl. fem./neut.

the nom.sg.f. *indí*, which is the Wb. 8a18 case and the two Sg. forms in (1) above. The spelling *intí* for the nom.sg.m. form, however, is consistent (I have not found spellings other than that), in spite of the minimal difference with respect to the feminine form *indí*, so that the scarcity of feminine forms cannot be attributed to a generalized orthographic or even formal confusion. This scarcity is linguistically real and is probably due to pragmatic and/or semantic reasons.

The two features of the general distribution of gender in the Old Irish light heads observed in this and in the previous sections, namely, the virtual lack of feminine forms and the somewhat more frequent character of the neuter forms in the corpus, are probably related to each other, in the sense that both seem to be due to the same generic character of the NPs introduced by the light heads. It is on the basis of this distribution that it makes sense to talk about a virtual opposition between animate and inanimate light heads, in the sense that the majority of masculine forms are actually animate and less so masculine in the strict sense of the term.

On most occasions, the animate light-headed NPs are humans, there being some exceptions such as example (1) above, in which the referent intended is 'grammatical person'. Old Irish has certainly a threefold gender distinction, but the light heads have a similar behavior to other elements such as the indefinite pronoun *nech ní* 'some-/anybody, some-/anything', which only distinguishes animate and inanimate.

6.2. Attestation types in the glosses and animacy

The distribution of the NPs introduced by animate and inanimate light heads according to the attestation types described in Section 3 above is first given in Table 6, which includes all forms of whichever case and number. This table is based on Table 5, and puts together the right- and left-dislocated NPs into a single group.

Table 6. Gender distribution and attestation types for all forms of the light heads *intí/aní* and a^N

Attestation type	Wb.			Ml.			Sg.			Total		
	intí	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N
	m./f.	n.		m./f.	n.		m./f.	n.		m./f.	n.	
Isolated NPs	24	9	13	74	17	7	4	68	0	102	94	20
Dislocated NPs	23	8	18	16	13	1	0	3	0	39	24	19
Intraclausal NPs	24	22	67	31	87	20	8	52	2	63	161	89
Total 1	71	39	98	121	117	28	12	123	2	204	279	128
Anim. vs. inanim.	71	137		121	145		12	125		204	407	
intí/aní vs a^N	110		98	238		28	135		2	483		128

As a general observation on Table 6, the animate light-headed NPs in Wb. are evenly distributed in the three main types of attestations considered, whereas these NPs are in Ml. much more frequent as isolated NPs. In these two collections of glosses, the neuter light heads (i.e. both *aní* and a^N) clearly prefer the intraclausal type.

As noted in Section 5 above, Sg. displays an unbalanced situation in which there are actually very few masc./fem. forms and most of the attested neuter light heads are used in the quotation of Latin (and Greek) words, a situation that requires a separate treatment.

While the count of all animate and inanimate forms of Table 6 is surely informative for the research question, one could argue that, since left-dislocated NPs are only in nominative case, the NPs introduced by the light heads *intí/aní* and a^N to be considered in the comparison with the other attestation types should be only those in nominative case. Recall that, in Table 4 above, quite a lot of the isolated NPs in Ml. are in grammatical cases other than nominative.

This is why Table 7 includes only the nominative forms for all attestation types, as a further distribution to be considered. This means that Table 7 includes neither the intraclausal NPs introduced by the light heads *aní* and a^N that express P NP_{mat} function, nor the right-dislocated NPs introduced by the same forms *aní* and a^N that may be assumed to correspond to an object of the main clause.[14]

Table 7. Gender distribution for the attestation types of nominative singular and plural forms of the light heads *intí/aní* and a^N

Attestation type	Wb.			Ml.			Sg.			Total		
	intí	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N	*intí*	*aní*	a^N
	m./f.	n.		m./f.	n.		m./f.	n.		m./f.	n.	
Isolated NPs	18	7	13	25	10	7	3	24	0	46	41	20
Dislocated NPs	22	8	15	15	12	1	0	3	0	37	23	19
Intraclausal NPs	9	14	32	15	26	14	4	7	1	28	47	47
Total 1	49	29	60	55	48	22	7	34	1	111	111	83
Anim. vs. inanim.	49	89		55	70		7	35		111	194	
intí/aní vs a^N	78		60	103		22	41		1	222		83

The figures in Table 7 that are closer to those in Table 6 are those of the dislocated NPs. This is due to the fact that most of these NPs are left-dislocated, which are all in nominative case. Note further that the attestation type in which there are greater numerical differences between Tables 6 and 7 are the intraclausal ones, because the number of forms in grammatical cases other than nominative is somewhat greater than in the case of the isolated NPs.

14 In practice, this affects three cases of a^N, namely Wb. 10b1, and Wb. 15a34 (2x).

Finally, another grouping that could be considered is the one that only includes the nominative singular forms, due count of the fact that the light head *a*^{N} is only singular. These figures can be found in Table 8, which repeats the figures for *a*^{N} in Table 7.

Table 8. Gender distribution for the attestation types of nominative singular forms of the light heads *intí/aní* and *a*^{N}

Attestation type	Wb.			Ml.			Sg.			Total		
	intí	*aní*	*a*^{N}	*intí*	*aní*	*a*^{N}	*intí*	*aní*	*a*^{N}	*intí*	*aní*	*a*^{N}
	m./f.	n.		m./f.	n.		m./f.	n.		m./f.	n.	
Isolated NPs	10	4	13	13	5	7	3	19	0	22	28	20
Dislocated NPs	13	8	15	13	10	1	0	2	0	26	20	19
Intraclausal NPs	7	14	32	11	22	14	4	7	1	22	43	47
Total 1	30	26	60	37	37	22	7	28	1	70	91	83
Anim. vs. inanim.	30	86		37	69		7	29		70	184	
intí/aní vs *a*^{N}	56		60	74		22	35		1	161		83

In a statistical analysis of the general correlation between the attestation types proposed in Section 4 and the divide between animate vs inanimate advocated in Section 6.1, it is safer to proceed with this sort of homogeneous purposive sampling, provided that the proposed samples make sense.[15] This is why I have presented not only the raw quantities of NPs introduced by all animate and inanimate light heads, as an overview of the bare opposition based on animacy that ignores any other condition, but also other, more reduced but homogeneous samples, limited to the light heads in nominative case, either singular and plural, or only singular.

7. Statistical analysis

The data sets considered in this section are based on the three samplings presented in Tables 6 to 8 above. Within each of them, I also consider two possible groupings of the attestation types, namely, (a) each of the three types by itself, and (b) the isolated and the dislocated NPs as one group of non-intraclausal NPs vs the intraclausal ones. Since this paper looks at the effect of syntactic NP extraclausality, see Section 8.2 below, and this includes both the isolated and the dislocated NPs, other logically possible groups are not considered.

Due to the clearly skewed use of the neuter light heads in Sg., the data of this collection are excluded from the statistical analysis of this section. The prevalent Sg. use of the neuter light heads, in particular of those of *intí/aní*, is due to their use

15 For this type of sampling, see the general paper by Etikan *et alii* (2016: 3).

in the quotation of Latin and Greek words. In this sense, the data sets considered are those of Wb. and Ml. individually and the sum of both. The combination of the three samplings (according to NP case, to attestation type and to text) brings about a total number of 18 samples that will be submitted to statistical analysis.

Tables 9a to 9c take the three attestation types as one set each (hence, df= 2), and each consider the three groups considered in the previous section. That is to say, Tables 9a to 9c are based on Tables 6 to 8 above respectively.

Table 9a. Animate vs inanimate according to the variable isolated, dislocated and intraclausal NP for all NPs introduced by the light heads *intí/aní* + a^N: statistical tests

Data set	χ^2 (df= 2)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [24, 22, 23, 26, 24, 89]	18.59	0.00009	0.299	0.00008
Ml. [74, 24, 16, 14, 31, 107]	65.87	4.97E–15	0.497	6.23E–16
Wb. + Ml. [98, 46, 39, 40, 55, 196]	83.94	5.93E–19	0.420	1.97E–19

Table 9b. Animate vs inanimate according to the variable isolated, dislocated and intraclausal NP for NPs introduced by *intí/aní* and a^N in nom.sg./pl. case: statistical tests

Data set	χ^2 (df= 2)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [18, 20, 22, 23, 9, 46]	14.66	0.0006	0.325	0.0004
Ml. [25, 17, 15, 13, 15, 40]	11.39	0.0033	0.301	0.0028
Wb. + Ml. [43, 37, 37, 36, 24, 86]	25.00	0.000003	0.308	0.000002

Table 9c. Animate vs inanimate according to the variable isolated, dislocated and intraclausal NP for NPs introduced by *intí/aní* and a^N in nom.sg.: statistical tests

Data set	χ^2 (df= 2)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [10, 17, 13, 23, 7, 46]	8.16	0.0169	0.265	0.0147
Ml. [13, 12, 13, 11, 11, 36]	8.93	0.0114	0.305	0.0113
Wb. + Ml. [23, 29, 26, 34, 18, 82]	16.22	0.0003	0.276	0.0002

Tables 10a to 10c make two sets, namely, 'non-intraclausal' (i.e. isolated + dislocated NPs) vs 'intraclausal' attestation types of the NPs introduced by the light heads (hence, df= 1), and are again based on Tables 6 to 8 respectively.

Table 10a. Animate vs inanimate according to the variable non-intraclausal (i.e. isolated + dislocated) vs intraclausal NPs for all NPs introduced by *intí/aní* and a^N: statistical tests

Data set	χ^2 (df= 1)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [47, 48, 24, 89]	17.07	0.00003	0.296	0.00003
Ml. [90, 38, 31, 107]	59.40	1.28E–14	0.480	2.59E–15
Wb. + Ml. [137, 86, 55, 196]	74.91	4.92E–18	0.401	1.09E–18

Table 10b. Animate vs inanimate according to the variable non-intraclausal (i.e. isolated + dislocated) vs intraclausal for the NPs introduced by nom.sg./pl. forms of *intí/aní* and a^N: statistical tests

Data set	χ^2 (df= 1)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [40, 43, 9, 46]	13.28	0.0002	0.325	0.0001
Ml. [40, 30, 15, 40]	9.97	0.0015	0.298	0.0010
Wb. + Ml. [80, 73, 24, 86]	23.59	0.000001	0.307	5.31E–7

Table 10c. Animate vs inanimate according to the variable non-intraclausal (i.e. isolated + dislocated) vs intraclausal NPs for the NPs introduced by nom.sg. forms of *intí/aní* and a^N: statistical tests

Data set	χ^2 (df= 1)	*p*-value	Cramér's V	Fischer's ex. test
Wb. [23, 40, 7, 46]	6.98	0.0082	0.265	0.0054
Ml. [26, 23, 11, 36]	7.70	0.0055	0.304	0.0034
Wb. + Ml. [49, 63, 18, 82]	15.04	0.0001	0.276	0.00005

None of statistical analyses in Tables 9a,b,c and 10a,b,c obtains a *p*-value above the usual threshold of 0.05. Wb. and Ml., taken in themselves or conjointly, point to the same conclusion, namely, that there is a significant correlation between animacy of the light-headed NP and the attestation type: Whereas the animate ones are evenly distributed along the three attestation types considered, if not inclined to appear more often as non-intraclausal, there is a clear preference of the inanimate light-headed NPs for the intraclausal use.

8. Discussion: syntactic freedom and animacy

8.1. A closer look at the Old Irish isolated light-headed NPs

As Tables 6 to 8 above show, the isolated light-headed NPs align with the corresponding dislocated NPs or even offer a negative mirror image of the corresponding intraclausal NPs from a quantitative point of view. This means that

these isolated NPs are similar to the dislocated ones, and this is a result that should be carefully considered, following the suggestion in Section 4 above on the possibility of borderline cases. This section deals with some other examples from the corpus that stand in a similar situation or that may receive a linguistic interpretation in terms of alone standing NPs, as elements that make sense from a linguistic viewpoint.

Certainly, isolated light-headed NPs are often simply the bare translation of a Latin expression, very often a participle agreeing with a noun, as a situation that is specially clear when it is in a grammatical case other than nominative. Since such a use of the Latin participles is alien to the Old Irish grammar, an explanatory gloss of this type would serve as an aid for the better understanding of the Latin text. Consider examples (9) and (10), in which the light-headed NPs have been classified as isolated.

(9) *innaní asidgrennat* (Ml. 18d2)
innaní as-id$^{(N)}$·grenn-at
LHD.GEN.PL.M PV-3SM.REL·persecute.PRES.IND-3PL.ACT
'of those that persecute him'.

(10) *am*al *inni asroilli* .i. *air atroilli dia a aigsin donaib doinib* (Ml. 51d12)
amal inni as·$^{(L)}$/roilli *.i.*
as LHD.ACC.SG.M PV·(REL)/deserve.PRES.IND.3SG.ACT i.e.
air a-t$^{(L)}$·roilli dia
for PV-3SN.DECL·deserve.PRES.IND.3SG.ACT God.NOM
a$^{(L)}$=aigs-in do-n-aib=doin-ib
3SM.POSS=fear-ACC to-ART-DAT.PL.M=man-DAT.PL
'as one who deserves, i.e. for God deserves it to be feared by men'.

Example (9) is added to the Latin participle *persequentium* 'of (his) prosecutors'. The light-headed NP in (10) is added to (and translates) Latin *tanquam benemerentem*. Note that the latter case is accompanied by a further comment in the gloss, though this does not change its classification as isolated.

There are other isolated light-headed NPs that are not used to translate Latin participles or similar structures, however.

To begin with, a given isolated light-headed NP could be interpreted as an intraclausal argument, as in the case of (11), in which the presence of *act* 'but' implies the presence of a previous context, the one involved in the Latin expression to which the gloss is attached: *quod nihil aliud sapietis* 'that you may not know any other thing'. As a further possibility, the case in (11), which is introduced by *.i.* 'i.e.', is added to the Latin expression *id quod ex uobis deerat erga meum obsequium* 'that which was lacking of you with respect to my service',

and may well be interpreted as a right-dislocated NP that aims at explaining more clearly a previously mentioned referent, as in the case of (4) above.

(11) .i. *act aní ropridchussa dúib* (Wb. 20b4)
.i. act aní ro·$^{(L)/}$pridch-us=sa
i.e. save LHD.ACC.SG.N PRF·(REL)/preach-PRET.1SG.ACT=EMPH(1SG)
dú-ib
to-2PL
'save what I have preached unto you'.

(12) .i. *aní rop*r*idchus* (Wb. 23d18)[16]
.i. aní ro·$^{(L)/}$pridch-us
i.e. LHD.NOM.SG.N PRF·(REL)/preach-PRET.1SG.ACT
'what I have preached'.

Still, despite the possibility of assuming a syntactic continuity of the Latin text, examples (11) and (12) are classified as isolated NPs.

Some other NPs classified as isolated may be interpreted more clearly as utterances on their own. In (13), the NP *indí ɔsechat hulcu* et*mórate mathi* 'they that correct the evil and magnify the good', with the nom.pl.m. form *indí*, is actually the answer to the question appearing previously in the gloss in which there is a plural interrogative pronoun (*citné*).

(13) *assindet sunt tra citné cumacte diandid cóir infognam* .i. *indí ɔsechat hulcu* et*mórate mathi* (Wb. 6a9)
ass·ind-et sunt tra citné
PV·(DECL)/declare.PRES.IND-3PL.ACT here then WH.NOM.PL
cumacte di-a^{N}-did·cóir
power.NOM.PL of-OBL.REL-COP.PRES.IND.3SG·proper.NOM.SG.M
in=fognam *.i.* indí
ART.NOM.SG.M=service.NOM i.e. LHD.NOM.PL.M

16 Following the suggestion by an anonymous reviewer on this example, I have decided to include the notation *.i.* 'i.e.' in all the examples quoted in this paper in which appeared at the beginning of the gloss. Actually, a good deal of the glosses of the corpus are introduced by this abbreviation, so that it is rather a graphical device indicating the beginning of the gloss, in line with its general function of explanatory text of whichever type (i.e. linguistic, notional, historical). This is why this gloss-initial '.i.' has not been included in the word count of the corpus offered in García-Castillero (to apper). However, it is true that this '.i.' does not appear systematically, so that there could be some specific reason for the general use and/or distribution of this abbreviation I'm not aware of. The presence of '.i.' makes no difference in the basic interpretation of the attestation types proposed for the light-headed NPs, unless otherwise proven, of course.

con·$^{(L)}$sech-at hulc-u *et*
PV·(REL)correct-3PL.ACT evil-ACC.PL and
móra-te math-i
magnify.PRES.IND-3PL.REL.ACT good-ACC.PL
'he [Paul] declares here, then, what are the powers unto which the service is proper, to wit, they that correct the evil and magnify the good'.

This 'responsive' use can be assumed for some cases of 'completely' isolated NPs introduced by light heads in the Old Irish glosses. Consider (14), which is attached to the Latin word *fenerator* 'money-lender' and can thus be understood as the answer to a question such as 'What is a *fenerator*?'.

(14) .i. *intí dianairlicther ni són* (Ml. 127a15)
.i. intí di-a^{N}·airlic-ther
i.e. LHD.NOM.SG.M to-OBL.REL·lend.PRES.IND-3SG.PASS
ni són
something/NOM ANA
'he to whom something is lent'.

Another worth-considering example is (15), which includes as much as three cases of light heads introduced by the nominative plural *indí* that have been classified as isolated NPs.

(15) .i. *duárchomraicset cloini n doib fesin* meriti .i. *indairílti* .i. *indí assidroilisset*· obtimi .i. *indfoircimi* [leg. *indfoircimim*]· de se · *díb* .i. *indí adidroillisset commór inclóini nísin dutairciud doib*· obtarent· .i. *indí assaguiset*· […] (Ml. 61b17)
.i. du·árchomraic-s-et cloini^{N} do-ib
i.e. PV·(DECL)/PRF.gather-PRET-3SG.ACT iniquity.ACC to-3PL
fesin *merit-i* *.i.* ind$^{(L)}$=airílti
REFL deserved-NOM.PL.M i.e. ART.NOM.PL.M=deserved.NOM.PL
.i. indí ass-id·roili-ss-et
i.e. LHD.NOM.PL.M PV-3SN.REL·deserve-PRET-3PL.ACT
obtim-i. *i.* ind$^{(L)}$=foircimim
best- NOM.PL.M i.e. ART.NOM.PL.M=best/SUPERL/NOM.PL.M
de *se* dí-b *.i.* indí
from REFL.ABL from-3PL i.e. LHD.NOM.PL.M
ad-id·roilli-ss-et co^{N}mór
PV-3SN.REL·deserve-PRET-3PL.ACT greatly
in=clóini/^{N} í=sin du-tairciud
ART.ACC.SG.N=iniquity/ACC DEM=DIST to-causing/DAT

do-ib *obt-are-nt* *.i.* indí
to-3PL wish-PRET.SUBJ-3PL.ACT i.e. LHD.NOM.PL.M
ass-a$^{(L)}$·guis-et
PV-REL·wish.PRES.IND-3PL.ACT
'they gathered iniquity to themselves; *meriti*, i.e. deserved, i.e. those who deserved it; *optimi*, i.e. the most excellent; *de se*, of them; i.e. they who greatly deserved that that iniquity should be caused to them; *optarent*, i.e. those who wish, […]'.

As for the light-headed NPs introduced in the quote, it is clear that they are not of the intraclausal type: apart from appearing after the abbreviation *.i.* 'i.e.', which conveys a prosodic and syntactic pause, in the specific context in which they appear, they certainly provide a translation of Latin forms, so that the first *indí assidroilisset* 'those who deserve it' somehow translates *meriti*, but the proper translation is the previous *ind-airilti*, and the light headed NP is a sort of additional comment, similar to a right-dislocated NP, but also similar to an answer of a question such as 'who are these *meriti*?'. The two other NPs introduced by the same form *indí* stand in a similar situation, in which almost whichever interpretation, namely, bare translation, right-dislocated NP, or even answer to an implicit question, would be valid. In my opinion, this uncertainty is a sign of the similar character of these three interpretations that have been illustrated in the previous examples in this section.

It is important to insist in that purely translating glosses such as the ones in (9) above are no more than that, i.e. that they may have no special linguistic significance. However, some other cases in which one may assume a translating function lend themselves to a properly 'linguistic' interpretation in terms of a right-dislocated NP or an answer. In fact, these two types of utterances are a sort of explanation, which is also the function of a translation.

8.2. The notions of syntactic NP independency and extraclausality

Since light headed NPs classified as 'isolated NPs' such as those in (13) and (14) may represent real utterances in themselves, and those in the gloss of (15) may be interpreted in terms of dislocated or isolated NPs, which points to a certain functional contiguity between these types, it is worth looking at the shared feature(s) of these NPs that involve extraclausality or, to state it with a perhaps wider notion, non-intraclausality. I think that the involved feature is 'syntactic independency', that I provisionally define in this case as the ability of a given NP to appear in a non-intraclausal environment.

The notion of syntactic independency has been suggested by some scholars, notably in Evans' (1993) seminal paper in which this notion is not only

considered for both NPs and (insubordinate) clauses, but also for the Welsh responsive forms. As also noted by Fox & Thompson (2010), an answer to a previous question is rather an utterance on its own in which one may assume some sort of ellipsis. The same phenomenon has been termed 'analepsis' by Auer (2014), NP 'insubordination' by Cristofaro (2016: 403–404), a sort of 'cooptation' by Heine *et al.* (2016: 44), or 'free NP' by Helasvuo (2019). For other similar analyses of dislocated and, in more general terms, extraclausal NPs, I refer to Progovac *et aliae* (2006), Giomi & Keizer (2019), and Villa-García (2023: 4).

A separate mention deserves Ono & Thompson's (1994) paper, which links isolated ('unattached') NPs with dislocation. These scholars offer a brief functional characterization of the 'unattached NPs' in conversational English, which would have two basic functions, referential and predicating. Whereas the referential function, according to Ono & Thompson (1994: 404, 407) is much the same as the one of the dislocated NPs, and this clearly agrees with the above observations on example (15), the latter is more similar to the function assumed for examples (11), (13) and (14) above. In particular for the predicating cases, Ono & Thompson (1994: 409) argue against the idea of a process of deletion or ellipsis.

In their analysis of the differences between both functions, Ono & Thompson (1994: 411) note that the predicating function is not expected for highly referential NPs such as names and pronouns, and this is relevant at this point because the notion of extraclausality and even isolation plays a decisive role in the distribution of the Old Irish pronominal forms.

For reasons of space, I refrain from dealing in any depth with the notion of NP extraclausality (but not necessarily isolation) in Old Irish, for which I have to refer to García-Castillero (2019) in general, and to García-Castillero (2013) for the Old Irish tonic pronouns. Suffice it to say that the Old Irish tonic pronouns, which are only used as extraclausal NPs, and the dislocated light-headed NPs stand in a sort of complementary distribution: the latter do not appear as focused constituents and fulfill the extraclausal function that is by far much less frequent in the former.

8.4. Animacy and NP syntactic independency

Animacy is a decisive factor in the use of some Old Irish elements such as the elusive *notae augentes*. Griffith (2008) has convincingly demonstrated that, in a verbal complex that includes a 3rd and a 1st or 2nd person pronominal referent, the *nota augens* selected corresponds to the pronominal reference which stands higher in the animacy hierarchy, namely, to the 1st or 2nd person.

The correlation between non-intraclausal (i.e. isolated and dislocated) light-headed NPs and animate referents, on the one hand, and intraclausal NPs and

inanimate ones, on the other, is not difficult to justify. If it is accepted that 'isolated NPs' are functionally close to the dislocated NPs, at least for the case of the light-headed NPs in Old Irish, one could assume that what applies to the dislocated NPs also applies to the isolated ones.

Thus, since inanimate referents tend to be objects more frequently than the animate ones,[17] they are expected to appear more often in the intraclausal domain in which the object appears. Conversely, since animate referents tend to be more frequently subjects, and less so objects, they may appear more easily in extraclausal domains in which they are also expected due to their inherent saliency, a cognitive property that is assumed for animates in psycholinguistic studies. The following quote, in which the elided parts are bibliographical references, is the general characterization provided by Vihman & Nelson (2019) in the introductory paper on a collective volume on the effects of animacy in grammar and cognition.

> […], animate beings are likely to be viewed as potential and actual agents, and to be expressed as subjects […]. The typical alignment between grammatical subjects, pragmatic topics and referential form […] can be seen to also support a strong tendency for subjects and topics to refer to animate beings, especially humans […]. Likewise, the preference to structure clauses according to a 'human first' principle […] can be seen to derive from the typical structure of events humans are interested in talking about. (Vihman & Nelson 2019: 261).

These expected trends fit in with the situation observed for the light-headed NPs in the Old Irish glosses, in particular, in Wb. and Ml.

In a study on the correlation between animacy and NP fronting in the English Switchboard corpus of telephonic conversations, Snider & Zaenen (2006: 333) observe that animates are more frequent than inanimates in left-dislocated NPs.

There are surely more arguments to be offered for the general correlation between extraclausality and animate NPs advocated in this section, in both the theoretical and evidential sides. I think, however, that the aforementioned observations suffice to accept the notion of 'NP syntactic independency', parallel to the notion of 'syntactic (in)dependency' used in the definition of clause subordination, as a useful theoretical tool to explain the observed tendencies for Old Irish animate and inanimate light-headed NPs.

---

[17] In the present case, and bearing in mind only the intraclausal light-headed NPs of the corpus, there are 4 instances with P NP_{mat} function out of 63 animate NPs (i.e. 6.35%), as against 66 out of 250 inanimate ones (i.e. 26.4%). The most frequent NP_{mat} function of the inanimate light-headed NPs in Old Irish is S, as noted in García-Castillero (in prep.), in agreement with the most frequent profile of S NP_{mat} and P NP_{rel} functions of these referents.

9. Conclusion

Osborne's (2013) conclusion regarding missing data acquires a special significance in corpus linguistics, at least in the case considered in this paper.

> Missing data is often viewed as lost, an unfilled gap, but […] it is not always completely lost, given the availability of other strongly correlated variables. Going one step farther, missingness itself can be considered as outcome itself, and in some cases can be an interesting variable to explore. There is information in missingness. (Osborne 2013: 128–129)

Furthermore, the case presented in this paper may well be a case of the so-called missing data not at random (MNAR), that Osborne (2013: 109) describes as implying that missingness is correlated to some variable. In the present case, the lack of NP_{mat} function of the isolated and dislocated light-headed NPs in the Old Irish glosses correlates with the animate character of the NP, whereas these attestation types are significantly less frequent with their inanimate counterparts, which are more often used as intraclausal NPs. This correlation agrees with general trends observed in the literature on the phenomenon of 'syntactic NP independency'.

The notion of missing data seems therefore to find a good accommodation in the field of modern corpus historical linguistics. For a study on the syntax of the light-headed NPs in the glosses, we do not need to completely discard all NPs that have no external syntax, as if the lack of NP_{mat} function would invalidate it for any consideration about their syntax.

The light-headed NPs constitute a relatively coherent and numerous group, but there are other NP types in the Old Irish glosses that would deserve a separate treatment. A brief reference has been made in Section 8.2 to stressed pronouns. In the investigation of NP syntax, in particular before trying to formulate principles of a general nature, it seems wise to proceed step by step with coherent groups, that may or may not fully agree with other NP types. A further, more general issue that also lies beyond the scope of this paper is the correlation of animacy and syntactic (in)dependency of NPs.

REFERENCES

Andrews, Avery D. 2007. "Relative clauses", in:Timothy Shopen (ed.), *Language Typology and Syntactic Description. Volume 2. Complex Constructions* (2nd ed.). Cambridge: Cambridge University Press. 206-236.

Auer, Peter. 2014. “Syntactic structures and their symbiotic guests. Notes on analepsis from the perspective of on-line syntax”, *Pragmatics* 24(3): 533-560.

Bauer, Bernhard and Stefan Schumacher. 2015. *The Online Database of the Old Irish Priscian Glosses*. Available at http://www.univie.ac.at/indogermanistik/priscian/.

Bisagni, Jacopo. 2013-2014. “Prolegomena to the study of code-switching in the Old Irish Glosses”, *Peritia* 24-25: 1-58.

Breatnach, Liam. 1990. “On the citation of words and a use of the neuter article in Old Irish”, *Ériu* 41: 95-101.

Breatnach, Liam. 2020. “The demonstrative pronouns in Old and Middle Irish”, in: Elliott Lash, Fangzhe Qiu, and David Stifter (eds.), *Morphosyntactic variation in Medieval Celtic languages (Corpus-Based Approaches)*. Berlin/Boston: De Gruyter Mouton, 115-141.

Citko, Barbara. 2004. “On headed, headless, and light headed relatives”, *Natural Language and Linguistic Theory* 22: 95-126.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

CorPH = Stifter *et alii* (2021).

Cristofaro, Sonia. 2016. “Routes to insubordination: A cross-linguistic perspective”, in: Nicholas Evans and Honoré Watanabe (eds.), *Insubordination*. Amsterdam and Philadelphia: John Benjamins, 393-422.

Diessel, Holger. 1999. *Demonstratives (Form, function and grammaticalization)*. Amsterdam and Philadelphia: John Benjamins.

eDIL 2019: *An Electronic Dictionary of the Irish Language*, based on the Contributions to a Dictionary of the Irish Language (Dublin: Royal Irish Academy, 1913-1976) (www.dil.ie 2019). Accessed on [2023-12-12].

Epps, Patience. 2009. “Escape from the noun phrase (From relative clause to converb and beyond in an Amazonian language)”, *Diachronica* 26(3): 287-318.

Etikan, Ilker, Sulaiman Abubakar Musa & Rukayya Sunusi Alkassim. 2016. “Comparison of Convenience Sampling and Purposive Sampling”, *American Journal of Theoretical and Applied Statistics* 5(1): 1-4.

Evans, Nicholas. 1993. “Code, inference, placedness and ellipsis”, in: William A. Foley (ed.), *The Role of Theory in Language Description*. Berlin and New York: Mouton de Gruyter, 243-280.

Fox, Barbara A. and Sandra A. Thompson. 2010. “Responses to Wh-Questions in English Conversation”. *Research on Language and Social Interaction* 43(2): 133-156.

García-Castillero, Carlos. 2013. “Old Irish tonic pronouns as extraclausal constituents”, *Ériu,* 63, 1-39.

García-Castillero, Carlos. 2018. “The name and its circumstances: on the recognitional use of the Old Irish demonstrative *intí*”, in: José M. Vallejo, Iván Igartua, and Carlos García-Castillero (eds.), *Studia Philologica et Diachronica in Honorem Joaquín Gorrochategui (Indoeuropaea et Palaeohispanica)*. Vitoria-Gasteiz: Servicio de Publicaciones de la Universidad del País Vasco, 147-169.

García-Castillero, Carlos. 2019. “Towards a grammar of parenthetical constituents: synchronic and diachronic aspects of the Old Irish extraclausal noun phrases”, *Sprachwissenschaft,* 44/2, 223-255.

García-Castillero, Carlos. 2020. *Clause Typing in the Old Irish Verbal Complex (Morphological expression, paradigmatic organization and syntactic implications)*. Berlin and Boston: De Gruyter-Mouton.

García-Castillero, Carlos. In preparation. *Studies in Old Irish Clause Subordination (A corpus-driven analysis of linguistic variation).*

Giomi, Riccardo and Evelien Keizer. 2020. "Extra-clausal constituents in Functional Discourse Grammar: Function and form", *Revista da Abralin* 19/3: 159-185.

Griffith, Aaron and David Stifter. 2013. Dictionary of the Old Irish Glosses in the Milan MS Ambr. C301 inf., available at ULR http://www.univie.ac.at/indogermanistik/download/Milan_Glosses/.

Griffith, Aaron. 2008. "The animacy hierarchy and the distribution of the *notae augentes* in Old Irish", *Ériu* 58: 55-75.

Griffith, Aaron and David Stifter. 2014. "New and corrected ms. readings in the Milan glosses", *Études Celtiques* 40: 53-84.

Helasvuo, Marja-Liisa. 2019. "Free NPs as units in Finnish", *Studies in Language* 43(2): 301-328.

Hofman, Rijcklof, Pádraic Moran & Bernhard Bauer, *St Gall Priscian Glosses*, version 2.1 (2023) <http://www.stgallpriscian.ie/forms> [accessed 15 July 2023].

Jenset, Gard B. and Barbara McGillivray. 2017. *Quantitative Historical Linguistics*. Oxford: Oxford University Press.

Kavanagh, Seamus. 2001. *A Lexicon of the Old Irish Glosses in the Würzburg Manuscript of the Epistles of St.Paul*, ed. by Dagmar S. Wodtko, Wien: Österreichische Akademie der Wissenschaften.

Lash, Elliott. 2011. *A Synchronic and Diachronic Analysis of Old Irish Copular Clauses*. Cambridge: Cambridge University dissertation.

Mantenuto, Iara and Ivano Caponigro. 2021. "Light-headed relative clauses in Teramano", *Studies in Language* 45(4): 921-937.

Ono, Tsuyoshi and Sandra A. Thompson. 1994. "Unattached NPs in English Conversation", in: *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society: General Session Dedicated to the Contributions of Charles J. Fillmore*. Berkeley: University of California Press, 402-419.

Osborne, Jason W. 2013. *Best Practices in Data Cleaning (A complete guide to everything you need to do before and after collecting your data)*. Thousand Oaks CA: Sage.

Progovac, Ljiljana, Kate Paesani, Eugenia Casielles & Ellen Barton (eds.). 2006. *The Syntax of Nonsententials (Multidisciplinary perspectives)*. Amsterdam and Philadelphia: John Benjamins.

Roma, Elisa. 2023. "Headless relative clauses in Early Irish", *Études Celtiques* 49: 115-134.

Snider, Neal & Annie Zaenen. 2006. "Animacy and Syntactic Structure: Fronted NPs in English", in: Miriam Butt, Mary Dalrymple and Trace Holloway (eds.), *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*. Stanford CA: CSLI Publications, 323-338

Stifter, David, Bernhard Bauer, Elliott Lash, Fangzhe Qiu, Nora White, Siobhán Barrett, Aaron Griffith, Romanas Bulatovas, Francesco Felici, Ellen Ganly, Truc Ha Nguyen and Lars Nooij. 2021. *Corpus PalaeoHibernicum* (= CorPH) v1.0. Online at http://chronhib.maynoothuniversity.ie.

Stokes, Whitley and John Strachan (eds.). 1901-1903. *Thesaurus Palaeohibernicus. A collection of Old Irish glosses, scholia, prose and verse*, (with suppl. by Whitley Stokes). Dublin: Dublin Institute of Advanced Studies.

Thurneysen, Rudolf. 1946. *A grammar of Old Irish.* Revised and enlarged edition. Translated from the German by Daniel A. Binchy and Osborn Bergin. Dublin: Dublin Institute of Advanced Studies (reprinted with supplement 1975).

Untermann, Jürgen. 1983. “Indogermanische Restsprachen als Gegenstand der Indogermanistik”, in: Edoardo Vineis (ed.), *Le lingue indoeuropee di frammentaria atestazione / Die indogermanischen Restsprachen (Atti del Convegno della Società Italiana di Glottologia e della Indogermanische Geschellschaft, Udine, 22-24 sett. 1981)*. Pisa: Giardini, 11-28.

Vihman, Virve-Anneli & Diane Nelson. 2019. “Effects of Animacy in Grammar and Cognition: Introduction to Special Issue”. *Open Linguistics* 5: 260-267.

Villa-García, Julio. 2023. “Hanging Topic Left Dislocations as extrasentential constituents: toward a paratactic account. Evidence from English and Spanish”. *The Linguistic Review*: 1-46.

Wiechmann, Daniel. 2015. *Understanding Relative Clauses (A Usage-Based View of the Processing of Complex Constructions)*. Berlin: De Gruyter & Mouton.