

A new perspective on pro drop in Romanian

Olaf Koeneman & Hedde Zeijlstra

Abstract

A strong correlation can be observed between the availability of rich agreement morphology in a language and the ability to leave out the subject from a tensed clause (cf. Taraldsen 1978, Rizzi 1982 and many others), but it has proven hard to understand this correlation in its details. In this paper, we will provide a new theory on pro drop, which holds that pro drop is only possible if a language expresses tense and agreement by separate morphemes (Koeneman & Zeijlstra, under review). We will consequently apply this theory to Icelandic and Romanian. Although these languages share important properties on the surface (they look equally rich), only the latter licenses argumental null subject. We will show how our proposal can correctly derive this difference.

Keywords: pro drop, null subjects, rich agreement, Romanian, Icelandic

1 Introduction

It is a well-known fact that a language like Italian can leave out a subject in a tensed clause (i.e., it licenses null subjects, or pro drop), whereas this leads to ungrammaticality in English. The contrast is shown in (1):

- (1) a. Gianni ha detto che ha telefonato Italian
Italian Gianni has said that has.3SG telephoned
'Gianni said that he called'
b. *John said that has telephoned English

This difference can be linked to the fact that Italian has a rich agreement system but English does not. As can be seen in (2), Italian shows six agreement distinctions in the present tense paradigm, whereas English only shows two:

- (2) Present and imperfect/past tenses in Italian and English

	ITALIAN		ENGLISH	
	amare ('to love')		to talk	
	<i>present</i>	<i>imperfect</i>	<i>present</i>	<i>past</i>
1SG	amo	amavo	talk	talked
2SG	ami	amavi	talk	talked
3SG	ama	amava	talks	talked
1PL	amiamo	amavamo	talk	talked
2PL	amate	amavate	talk	talked
3PL	amano	amavano	talk	talked

The link between rich agreement and null subjects makes a lot of intuitive sense: if the agreement form expresses (almost) the same information as the subject does, the latter can

easily go unexpressed. There is, however, a fundamental issue with this idea. The 3rd person singular form in English is as unique to the English agreement system as the 3rd person singular form in Italian is. This raises the question why English does not allow a null in at least 3SG contexts.

The usual solution is to formulate a binary parameter and assume that a negative setting (a consequence of poor agreement under some definition) disallows null subjects across the board, whereas a positive setting (a consequence of rich agreement under some definition) will allow the language to generate null subjects across the board. The problem for such a parametric approach, however, is the existence of so-called partial pro drop languages, which allow null subjects in some but not all contexts. Frisian, for instance, allows null subjects in 2SG contexts but not elsewhere (cf. De Haan 1984). Bavarian dialects are similar. Although some varieties also allow null subjects in 1PL and 2PL contexts, no variety allows them across the board (cf. Bayer 1984). Partial pro drop languages therefore show that licensing null subjects does not have to be an all-or-nothing affair and can be determined per (person/number) context. If so, why does English not allow it in 3SG contexts?

This problem aggravates when one considers a richer non-pro drop language like Icelandic, which has five distinctions in its agreement paradigm (cf. (3)). Despite the fact that it consequently marks four contexts uniquely, null argumental subjects are not licensed. This is especially noteworthy in the face of Romanian, which also shows five distinctions (shown for the first conjugation in (3)).

(3) Present and past/imperfect tenses in Icelandic and Romanian

	ICELANDIC		ROMANIAN	
	heyra ('to hear')		cânta ('to walk')	
	<i>present</i>	<i>past</i>	<i>present</i>	<i>past (imperfect)</i>
1SG	heyr-i	heyrði	cânt	cântam
2SG	heyr-ir	heyrðir	cânți	cântai
3SG	heyr-ir	heyrði	cântă	cânta
1PL	heyr-jum	heyrðum	cântăm	cântam
2PL	heyr-ið	heyrðuð	cântați	cântați
3PL	heyr-a	heyrðu	cântă	cântau

In contrast to Icelandic, Romanian is a full pro drop language like Italian. Now, it can be observed that the two contexts that in Romanian are syncretic in the present tense (3SG-3PL) are no longer syncretic in the imperfect tense, so one could perhaps argue that this is important for the licensing of null subjects. Note, however, that the exact same is true for Icelandic (cf. (3)). Given these similarities, it is hard to understand why Romanian is a full-fledged pro drop language, whereas Icelandic does not allow argumental null subjects at all.

In the next section, we will present a new proposal on how the link between agreement and pro drop is to be understood. We will argue that a language like Italian expresses tense and agreement in a transparent way, whereas this is not the case in languages like English and Icelandic. With this proposal in place, section 3 will look at Romanian in detail and show that it patterns with Italian and Spanish rather than with English and Icelandic. Section 4 deals with a central issue that Romanian poses for our proposal, having to do with an agreement alternation in the 3rd person singular. Two consequences of our analysis for understanding pro drop in Romanian are discussed in section 5, and section 6 concludes.

2 Mono- versus bi-morphemic expression of tense and agreement

We generally agree with the literature stating that rich agreement is a prerequisite for the licensing of pro drop.¹ Given the existence of the partial pro drop phenomenon, we assume that null subjects are in principle licensed contextually: if the agreement information on the verb is identical to the information expressed by a pronominal subject, the subject can go unexpressed.² These two assumptions together clearly overgenerate. They predict that any context that is uniquely marked (e.g. 3SG contexts in English and Icelandic, as well as all plural contexts in Icelandic) license null subjects, contrary to fact. There must therefore be another prerequisite for such licensing, which apparently Italian, Spanish and Romanian meet but English and Icelandic do not. This prerequisite, we propose, is the transparent marking of tense and agreement. What we mean by this is the following.

If you look at the present and imperfect tense systems of Italian and Spanish, we can clearly identify the past tense markers as *-v-* and *-b-* respectively. The agreement forms we discern in the present tense follow this past tense marker in the imperfect. This means that the forms at the surface provide strong evidence for the existence of two separate morphemes underlying these forms, one expressing tense features and another expressing agreement features. This is in contrast to English, where we observe that the only visible agreement form, the 3SG *-s*, does not return in the past tense.

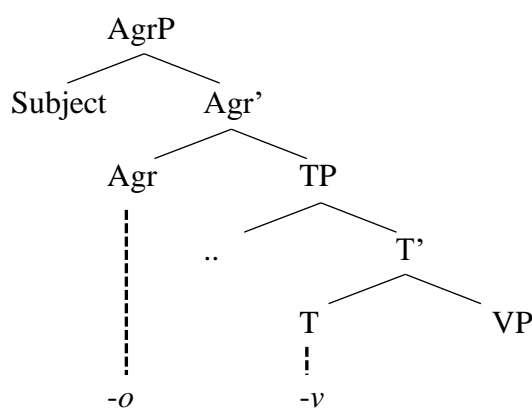
(4) Present and past/imperfect tenses in Italian, Spanish and English

	ITALIAN		SPANISH		ENGLISH	
	<i>present</i>	<i>imperfect</i>	<i>present</i>	<i>imperfect</i>	<i>present</i>	<i>past</i>
1SG	amo	amavo	amo	amaba	talk	talked
2SG	ami	amavi	amas	amabas	talk	talked
3SG	ama	amava	ama	amaba	talks	talked
1PL	amiamo	amavamo	amamos	amábamos	talk	talked
2PL	amate	amavate	amáis	amabais	talk	talked
3PL	amano	amavano	aman	amaban	talk	talked

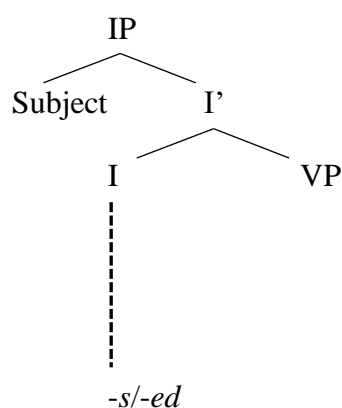
This has led Bobaljik & Thráinsson (1998) to conclude that English does not have separate morphemes for tense and agreement but a single one expressing both properties, I(NFL). In such an analysis *-s* and *-ed* are in direct competition for insertion into this one morphemic slot. The difference between Italian and English would then boil down to different settings of the so-called Split-IP Parameter: whereas English syntax generates INFL, Italian generates separate slots for tense and agreement. This is illustrated in (5) on the next page.

Given this background, we can now hypothesize that the difference between a bi-morphemic and a mono-morphemic expression of tense and agreement also underlies the difference between having or not having pro drop (see Saab 2008 for a similar idea). More specifically, we can say that English does not license null subjects because even in the context in which the agreement form is a unique marker (i.e., 3SG contexts) the morpheme expressing the relevant agreement features also expresses tense features. This morpheme is therefore overspecified, as it crucially expresses features (namely tense features) that are semantically incompatible with the subject that it is supposed to license. For null subjects to be possible, then, a split IP and a separate agreement morpheme constitute the relevant prerequisite.

(5) a. ITALIAN SPLIT-IP



b. ENGLISH UNSPLIT-IP



When we now turn to Icelandic, it is not immediately obvious how the analysis makes the right prediction. Take a look at (3) again. We could identify $-\ð-$ as the past tense marker that is followed by agreement forms, and these forms at least partially resemble forms we also see in the present tense. In short, Icelandic at first view looks like a split-IP language, and this is in fact how Bobaljik & Thráinsson analyze it. Of course, given our hypothesis, this would erroneously predict that Icelandic licenses argumental null subjects, contrary to fact.

However, there is one property that Icelandic shares with English: the form that appears in the 3SG present tense does not return in the 3SG past tense.³ English has $-s$ in the present tense but no overt 3SG agreement marker in the past. Icelandic has $-ir$ as the 3SG marker in the present tense, but $-i$ in the past tense. In other words, there is a morpho-phonological contrast in these 3SG contexts and this is in contrast with for instance Italian and Spanish. As can be observed in (4), the form in the 3SG context in the present tense is $(am)a$. The $-a$ following the stem can be treated as the theme vowel that we also see in the infinitive ($amare$ and $amar$) and in the plural contexts (1PL $amiano/amamos$, 2PL $amate/amáis$, 3PL $amano/aman$).⁴ Therefore, the 3SG agreement marker is null, contrasting with $-o$ and $-i/-as$ in the 1SG and 2SG, respectively. Crucially, this null form returns in the imperfect tense, where we see the exact same contrasts (we will return to the unexpected $-a$ in 1SG imperfect contexts in Spanish in section 3). In contrast to Icelandic and English, then, Italian and Spanish show no difference between the forms occurring in 3SG present and imperfect contexts. This particular property that sets Germanic agreement systems apart makes it harder to maintain a transparent analysis for tense and agreement, as we will now see.⁵

Let us look in more detail at the rules that insert concrete forms into the morphemic slots. We will make the standard assumption that the third person is the non-person in the sense of Forschheimer (1953) and Benveniste (1971) (see also Harley & Ritter 2002 and Preminger 2014). This entails that the form appearing in the 3SG context will function as the underspecified elsewhere form of the paradigm, expressing no feature values. Now, let us aim for a bi-morphemic analysis of Icelandic, starting with the following insertion rules for the singular part of the Icelandic paradigm:

- (6) $-i$ \triangleleft [Agr: speaker]
 $-ir$ \triangleleft [Agr:]

What the grammar of Icelandic now needs to capture is the following two properties: (i) 3SG $-ir$ becomes $-i$ in the past tense; (ii) $-ir$ remains $-ir$ in the 2SG past context. We first postulate a separate entry for the past tense marker $-\ð$.

(7) $-\delta$ $\langle \rangle$ [T: past]

Property (i) can then be achieved by postulating a context-sensitive rule which spells out the same features as *-ir* does in the present tense but inserts a different form. Such rules are needed generally to account for alternations between paradigms. One example is the 1PL context in Italian (cf. (4)), where the 1PL marker is *-iamo* in the present tense and *-amo* in the imperfect. This could be captured by adopting the following two rules:⁶

(8) $-iamo$ $\langle \rangle$ [Agr: speaker, plural]
 $-amo$ $\langle \rangle$ [Agr: speaker, plural] / [T: past]

Along the same lines, we can now postulate a context-sensitive rule for Icelandic to capture the fact that 3SG *-ir* in the present tense corresponds to *-i* in the past tense:

(9) $-i$ $\langle \rangle$ [Agr:] / [T: past]

Note, however, that under this analysis the *-i* appearing in 3SG past tense contexts is different from the *-i* appearing in 1SG past tense contexts: (9) must be assumed alongside the spell-out rule referring to [speaker] in (6). Therefore, the analysis postulates that these two instances of *-i* are accidentally homonymous instead of syncretic: it fails to capture their similarity directly. At the same time, the analysis now predicts that *-ir* becomes *-i* in 2SG past contexts too. In other words, it fails to capture property (ii). In order to prevent (9) from generating *-i* in 2SG past tense contexts, we have to assume that the *-ir* forms appearing in 2SG and 3SG present tense contexts are homonymous too and that the rule in (9) only competes with the one in (10)b.

(10) a. $-ir$ $\langle \rangle$ [Agr: addressee]
 b. $-ir$ $\langle \rangle$ [Agr:]

We conclude that a transparent, bi-morphemic analysis of tense and agreement can be maintained for Icelandic but at the cost of creating two homonym pairs.

Similar but different issues arise for English. In order to account for the fact that the 3SG *-s* does not return in the past tense, it must be assumed that *-s* competes with a null allomorph functioning as the elsewhere form in the past tense:

(11) $-\emptyset$ $\langle \rangle$ [Agr:] / [T: past]

This null allomorph will in addition be homonymous with the null form appearing in the 1SG and 2SG present and past contexts:

(12) $-\emptyset$ $\langle \rangle$ [Agr: participant]

We conclude that a bi-morphemic analysis can be maintained for English too, but at the cost of postulating a null allomorph and a homonymous pair.

There is an alternative analysis accounting for the absence of the 3SG present tense form in the past, namely one that assumes that the 3SG *-ir* form in Icelandic does not return in the 3SG past context because it directly competes with *-ði*. This analysis treats *-ði* as one form rather than two and requires the existence of an underlying morpheme expressing both tense and agreement features. In other words, the analysis treats Icelandic as a non-split-IP language. In English, *-s* would directly compete with *-ed* and this boils down to the analysis proposed by Bobaljik & Thráinsson (1998). Note that such an analysis does not have to postulate (11)

because *-s* is specified for present tense (cf. (13)). In addition, the mono-morphemic analysis also does not need to postulate a separate null form that spells out present tense (because the forms- \emptyset and *-s* in (13) already do that), thereby economizing on two null forms in total.⁷ The overall advantage of the mono-morphemic analyses is that they do not postulate unwanted null forms and homonyms. The singular parts of the paradigms can be captured by the following rules.

(13)	Icelandic		English
	<i>-i</i>	⟨ [I: present, speaker] ⟩	\emptyset
	<i>-ir</i>	⟨ [I: present, ___] ⟩	<i>-s</i>
	<i>-ði</i>	⟨ [I: past, ___] ⟩	<i>-ed</i>
	<i>-ðir</i>	⟨ [I: past, addressee] ⟩	

In sum, we have briefly explored the bi-morphemic and mono-morphemic takes on Icelandic and English and shown that in principle both are possible. From a theoretical perspective, it may be hard, or random, to choose one analysis over the other. The mono-morphemic analyses may economize on null allomorphs and homonyms, but only the bi-morphemic analyses directly capture the fact that all past forms contain *-ð* and *-ed*. After all, these forms are not listed as separate [T: past]-markers in the mono-morphemic analyses. If we reason from the perspective of child language acquisition, however, it becomes clearer why the mono-morphemic analyses have a plausible appeal. Take English, where the child observes that *-s* and *-ed* are in complementary distribution. To capture this, the child can either assume direct competition between the two overt form that (s)he has direct evidence for (namely *-s* and *-ed*) and settle for a mono-morphemic analysis, or create a competition effect between *-s* and an invisible form, the null allomorph in (11), for which there is no direct evidence. For the language learning child, the mono-morphemic analysis is therefore the more obvious choice as it relies on direct evidence. In the same vein, the mono-morphemic analysis is a way for the Icelandic child to not have to postulate superfluous homonymous pairs.

We can summarize our proposal in the following way:

- (14) If forms *x* and *y* are in complementary distribution, an analysis that takes *x* and *y* to be in direct competition takes precedence over an analysis that accounts for the effect by postulating a form *z*, where *z* is either a null form or a homonym.

It is important to understand that this proposal does not put any ban on null forms or homonym in general. They can be part of the child's analysis but are dispreferred if they are postulated as a way to capture a complementary distribution effect *between two other forms*. In that event, there is always an alternative analysis available, namely one that assumes direct competition between the two forms. Any null form or homonym that is not postulated to capture a complementary distribution effect of two other forms is therefore allowed, as (14) has nothing to say about these. To make this concrete, let us look at look at an unproblematic null form and homonym and contrast them with the problematic ones.

As an example of an unproblematic null form, take for instance the Italian paradigm in (4). Italian is like any other language discussed here in not expressing the present tense with an overt marker. Postulation of a non-overt one is straightforward, as such a null form stands in direct contrast with the *-v* that marks the imperfect. This null marker for the present tense is paradigmatically licensed and adopted as the spell out of present tense, which no other form in the paradigm would otherwise take care of. The status of this null form is therefore fundamentally different from the one that is made problematic by (14). The competing *x* and *y* forms in English are *-s* and *-ed*, respectively, and capturing their complementary distribution

by postulating an additional null form is dispreferred, given (14), and therefore problematic. In Italian, however, the *x* and *y* forms in competition are the present tense marker $-\emptyset$ and the imperfect marker $-v$. The statement in (14) does not even bear on this null form, as it simply does not rule out that either *x* or *y* itself is a null form.

A similar point can be made for homonymy. If there are two identical forms appearing in different slots of the paradigm, these forms will be analyzed as syncretic if the contexts in which we observe them shares features. Take as an example the Icelandic past tense:

(15) Icelandic past tense and German present tense

	ICELANDIC		GERMAN	
	heyra ('to hear')		spielen ('to play')	
	<i>past</i>		<i>present</i>	
	SG	PL	SG	PL
1	heyrði	heyrðum	spiele	spielen
2	heyrðir	heyrðuð	spielst	spielt
3	heyrði	heyrðu	spielt	spielen

In Icelandic, we can observe that the 1SG and 3SG form is identical. Since these appear in the same column, they obviously share a feature, namely the number feature [singular]. It is possible, therefore, to treat these two forms as syncretic. An analysis that does not do so and postulates two $-i$ (or $-\partial i$) forms instead assumes the existence of a homonymous pair that can in principle be avoided. Now, we saw in the analysis of Icelandic that there is a reason for postulating two $-i$ forms, namely to capture the absence of the 3SG present tense form $-ir$ in the 3SG past context under a bi-morphemic analysis. Since this essentially involves a complementary distribution effect of two forms, $-ir$ and $-\partial i$, (14) applies and the analysis is dispreferred. This Icelandic example stands in contrast to the German example in (15). Note for this language that the 3SG and 2PL contexts share the same form, namely $-t$, but these contexts are not in the same column, nor in the same row. In other words, they do not share a person or number feature. Therefore, postulating a homonymous pair of $-t$ forms is unproblematic for lack of a better analysis.⁸ Again, the statement in (14) does not even bear on this case, as the homonymous $-t$ pair is not postulated so as to capture a complementary distribution effect of two other forms.

This distinction between avoidable and non-avoidable null forms and homonyms becomes important when we turn to Romanian.

To conclude, Icelandic and English can be analyzed bi-morphemically and mono-morphemically but from the view of a language learning child there are clear reasons to prefer mono-morphemic ones for these languages, which we have formulated in (14). We can now turn to the analysis of Romanian. Recall that this language shares with Icelandic the number of agreement contrasts in the present tense and the fact that contexts that are syncretic in the present tense are no longer so in the past tense (cf. (3)). Since Romanian is a full-fledged pro drop language and Icelandic is not, the question is whether a bi-morphemic analysis can be maintained for Romanian. This will be explored in the next section.

3 A morphological analysis of Romanian

It is beyond the scope of this paper to provide an exhaustive analysis of the Romanian inflectional system. Below, we will focus on the agreement forms in the present and imperfect

tense of two different conjugations, the first and the fourth one. These show a different distribution of distinct forms, representative of the patterns found in Romanian at large. They constitute paradigms are fully productive and will therefore provide the child with the necessary information to choose between a split- or unsplit-IP analysis for the language. Wherever relevant, we will refer to other conjugations.

Let us start by looking at the two conjugations together.⁹ This leads to two important observations. The first one is that we can identify *-a* as the imperfect marker contrasting with no overt tense marker for the present tense. Since *-a* occurs in all contexts of the imperfect and irrespective of conjugation, it cannot be analyzed as a conjugation-specific theme vowel. The second important observation is that different conjugations have different agreement systems. For the first conjugation, it can be observed that each slot in the paradigm has its own agreement marker with the exception of the 3SG and 3PL forms, which are the same. If we look at the imperfect, however, we first of all see that the forms for 3SG and 3PL are distinct again (*-a* versus *-au*). In addition, we see that the 1SG and 1PL forms now look identical, in contrast what happens in the present tense. In the fourth conjugation, we observe a different pattern. In the present tense, the 3PL form is not identical to the 3SG form but to the 1SG form. In the imperfect tense, however, the 1SG and 3PL contexts show different agreement forms (*-am* versus *-au*), whereas the 1SG and 1PL are similar again, just like in the first conjugation.

(16) Romanian present and imperfect agreement paradigms

	FIRST CONJUGATION		FOURTH CONJUGATION	
	a cânta ('to sing')		a fugi ('to run')	
	<i>present</i>	<i>past (imperfect)</i>	<i>present</i>	<i>past (imperfect)</i>
1SG	cânt	cântam	fug	fugeam
2SG	cânți	cântai	fugi	fugeai
3SG	cântă	cânta	fuge	fugea
1PL	cântăm	cântam	fugim	fugeam
2PL	cântați	cântați	fugiți	fugeați
3PL	cântă	cântau	fug	fugeau

The question is whether we can provide an analysis for Romanian that captures these agreement alternations between present and imperfect tense in a bi-morphemic system (i.e., with all insertion rules either referring to tense or agreement features but not to a mixture of these two) while at the same time refrain from postulating avoidable null forms and homonym pairs, as these properties can trigger a switch to a mono-morphemic system, as we have hypothesized for Icelandic and English. We will first provide a general analysis of the first and fourth conjugations. These analyses will treat Romanian as a bi-morphemic language and will capture the basic facts. After that, we will focus on one problem for this analysis: as can be observed in (16), the 3SG form (i.e., *-ă* in the first and *-e* in the fourth conjugation) does not seem to return in the 3SG imperfect context. This initially suggests that Romanian runs into the same trap as English and Icelandic, raising the question why Romanian is not a mono-morphemic language without argumental pro drop. We will then sketch the two logical solutions available to us and show that the most plausible analysis is one in which the 3SG markers are in fact generated in imperfect contexts but targeted by phonological deletion, a process that can be independently justified for Romanian.

Let us start with an analysis of the first conjugation. Under the assumption that *-a* is the imperfect marker, the spell-out rules for tense are straightforward:

- (17) *Tense*
 -∅ <> [T:] (present)
 -a <> [T: past]

Like for any other language discussed here, the null form spelling out present tense is unproblematic. There is no other form spelling out this feature (it is not a null allomorph) and the null form stands in direct contrast with the *-a* marking the imperfect.

The spell-out rules for the agreement forms are less straightforward and require more discussion. We propose the following spell-out rules:¹⁰

- | | | | |
|------|------------------|---|---------------------------------------|
| (18) | <i>Agreement</i> | | <i>Tense</i> |
| | -∅ <> | [Agr: speaker, singular] | -∅ <> [T:] (present) |
| | -i <> | [Agr: addressee] | -a <> [T: past _{IMPERFECT}] |
| | -ă <> | [Agr:] | |
| | -m <> | [Agr: speaker] | <i>Theme vowels</i> |
| | -î <> | [Agr: addressee, plural] | -a <> [Th] / Conjugation 1 |
| | -u <> | [Agr: plural] / [T: past _{IMPERFECT}] | -ă <> [Th] / Conjugation 1, _[m] |

In order to account for the same form appearing in the 3SG and 3PL present tense contexts, we treat this similarity as syncretic by taking the form inserted in these contexts as the elsewhere form, which expresses no feature values. In the other slots, specific forms appear, and this is captured by the unique feature values that each form spells out: *-i* spells out [Agr: addressee] and *-î* [Agr: addressee, plural], etcetera. The theme vowel of the first conjugation is *-a* but in the present tense only occurs in 1PL and 2PL contexts.¹¹ Allomorphic theme vowel *-ă* occurs in 1PL contexts, presumably related to the phonological properties of the nasal agreement ending *-m*.

If we now turn to the imperfect tense, an issue appears: the 1SG and 1PL contexts show the same form, in contrast to what we observe in the present tense. In order to account for the fact that 1SG and 1PL contexts share the same form in the imperfect but not the present tense, we postulate *-m* as the form that spells out [speaker] but not a number feature. That allows us to treat the form in these two contexts as syncretic. However, we must also account for the fact that in the present tense 1SG contexts do not have *-m* but a null form. We therefore postulate a form *-∅* that spells out [speaker, singular]. Since the features underlying this form are more specific than the ones underlying *-m*, it blocks the latter from being inserted. In order to account for the fact that *-m* replaces *-∅* in the 1SG imperfect context, we propose that the [singular] feature is impoverished in the past tense by the impoverishment rule in (19):

- (19) [Agr: speaker, singular] → [Agr: speaker] / [T: past]

This blocks *-∅* from being insertion and *-m* is inserted instead as the less specified form compatible with that context.

Like context-sensitive spell-out rules, impoverishment rules (cf. Bonet 1991) are devices that can be used to account for agreement alternations between different parts of the paradigm. Specific for impoverishment rules is that they apply in marked parts of the paradigm (such as the plural or past tense dimension, cf. Aalberse & Don 2007) and lead to the appearance of unmarked forms over marked forms in those contexts. In other words, they capture the fact that another form already part of the paradigm fills the place of the disappearing form. Another example from Romance would be the Spanish imperfect, where we see that the contrast between 1SG and 3SG in the present tense disappears in the imperfect:

(20) Spanish agreement in present and imperfect tense

SPANISH		
amar ('to love')		
	<i>present</i>	<i>imperfect</i>
1SG	amo	amaba
2SG	amas	amabas
3SG	ama	amaba

We can account for this pattern in a similar way as we dealt with the pattern in Romanian: by impoverishment of a feature that is linked to the more specific form, in this case *-o*. The partial grammar (which treats all /a/-vowels following the stem *am-* as theme vowels, cf. Oltra-Massuet 1999) would look as follows:

(21)	<i>Agreement</i>		<i>Impoverishment</i>
	<i>-o</i>	⟨ [Agr: speaker]	[Agr: speaker] → [Agr:]/[T: past]
	<i>-s</i>	⟨ [Agr: addressee]	
	<i>-∅</i>	⟨ [Agr:]	

With impoverishment rules as a device to account for agreement alternations, an obvious question arises: what blocks us from using it to account for the disappearance of the present tense 3SG forms in the past tense in languages like English and Icelandic? This could perhaps allow us to uphold a bi-morphemic analysis for these languages without having to postulate avoidable null forms and homonyms. This would be an unwanted possibility, as the consequence would be that English and Icelandic are predicted to allow pro drop in at least those contexts that have a unique agreement marker.

The reason for the unavailability of impoverishment for English and Icelandic is as follows. If we look at Romanian and Spanish, we see that the alternations take place in 1SG contexts. These are contexts that express the [speaker] feature and in Romanian also the [singular] feature. These features can subsequently be targeted by impoverishment and the consequence is that a less marked form (the 1st person form in Romanian, and the 3SG elsewhere form in Spanish) will appear in that slot instead. When we look at Icelandic and English, however, we see that the alternation takes place in the 3SG context. This means that impoverishment should target any feature that is linked to this context. However, it is precisely the 3SG context which is the unmarked context expressing no feature values.¹² Hence, there is nothing that impoverishment could target. This means that impoverishment rules are a useless device to account for agreement alternations happening in 3SG contexts and the only remaining device is a context-sensitive rule. Such a rule must then postulate a null allomorph for English and a homonymous form for Icelandic, and this raises the acquisitional conundrum that we discussed.

Another property of the proposed analysis that is worth pointing out is the fact that in the present tense there is no dedicated form for the 3PL context. This is why the elsewhere form *-ã* also surfaces there. In the imperfect, however, there is a dedicated form, *-u*, which must be inserted by a context-sensitive rule. By restricting insertion of *-u* to imperfect contexts, we ensure that it is not inserted in 3PL present tense contexts.

Let us now turn to the fourth conjugation. Many aspects of our analysis of the first conjugation carry over to our analysis of the fourth one, but there are some specific details we have to focus on. We propose the following spell-out rules:

(22)	<i>Agreement</i>		<i>Tense</i>	
	-∅	⟨ [Agr: speaker, singular]	-∅	⟨ [T:] (present)
	-i	⟨ [Agr: addressee]	-a	⟨ [T: past _{IMPERFECT}]
	-e	⟨ [Agr:]		
	-m	⟨ [Agr: speaker]	<i>Theme vowel</i>	
	-îi	⟨ [Agr: addressee, plural]	-i	⟨ [Th] / Conjugation 4
	-∅	⟨ [Agr: plural] / Conjugation 4		
	-u	⟨ [Agr: plural] / [T: past _{IMPERFECT}]		

The main difference between the agreement systems of the first and fourth conjugations is the distribution of similar forms. Whereas in the first conjugation identical forms occur in 3SG and 3PL contexts, in the fourth conjugation identical forms appear in 1SG and 3PL contexts, which both display a null form. Since these two contexts do not form a natural class, the only way to analyze the null form as syncretic is by taking it to be the elsewhere form. We have assumed throughout, however, that the form occurring in the 3SG context always functions as the elsewhere, and this assumption was part of the explanation for why languages like English and Icelandic switch to a mono-morphemic analysis of tense and agreement: an elsewhere form cannot be blocked by impoverishing the elsewhere morpheme that it spells out, because that morpheme does not encode any features that can be targeted by impoverishment. We are therefore bound to the assumption that the 3SG form is the elsewhere form. Abandoning it would allow us to capture the similar forms occurring in 1SG and 3PL contexts as syncretic, but we would lose the explanation for why Icelandic (and English) do not allow argumental pro drop.

The consequence for the Romanian fourth conjugation is that we have to postulate two distinct null forms and the question arises if this is not problematic for us. The answer is no, for three reasons. First of all, since the two contexts displaying the null forms do not form a natural class (and the 3SG context is not one of the contexts in which the null form appears, excluding the analysis of it being an elsewhere form), this counts as a case of unavoidable homonymy. Moreover, both null forms are paradigmatically licensed, the first one through the contrast with the two other singular forms (2SG *-i* and 3SG *-e*), the second one through a contrast with the other plural forms (1PL *-m* and 2PL *-îi*). Finally, neither null form is postulated as a way to capture a complementary distribution effect between two other forms. In the opposition of 1SG *-∅* and 2SG *-i*, for instance, the 1SG *-∅* form is the x in (14) and the 2SG *-i* form is the y. As stated before, (14) does not rule out that either x or y is a null form itself. Therefore, (14) has nothing to say about the null forms postulated in (22). They are therefore as innocent as postulating a null form expressing present tense.

4 The problem of the 3SG forms

Let us now zoom in to a central problem of the Romanian paradigms, the observation that the form appearing in the 3SG of the present tense context does not return in the 3SG imperfect context. If imperfect tense is marked by *-a* and in the first conjugation the elsewhere agreement form *-ă* is inserted in 3SG imperfect contexts as well, we expect the outcome to be *cântaă* instead of *cânta*. For this reason, the 3SG person in the imperfect is sometimes taken as marked by a null form (cf. Zafiu 2012:32). This form then functions as a null allomorph, as it replaces the *-ă* form we see in the 3SG present tense. Given (14), the prediction is then that a learner of Romanian would switch to a mono-morphemic analysis, like the English and Icelandic learner,

with the unwanted consequence that we no longer expect pro drop to be possible. The question is therefore whether it is possible to rescue the bi-morphemic analysis.

There are two logical possibilities we can pursue. One is to say that the forms appearing in the 3SG present tense contexts are in fact generated in the imperfect as well but that they get phonologically deleted. Let's call this *the deletion account*. In that way, we do not have to postulate a null allomorph and Romanian retains its bi-morphemic status. A second option is to say that the *-ă* and *-e* forms we see in the 3SG present tense contexts are not agreement markers but present tense indicative markers (Illiescu & Mourin 1991). If that is the case, we can take the 3rd person agreement marker to be null, like in Italian and Spanish, and this null form can then be taken to transparently return in the imperfect tense. In that case, there is no reason anymore to consider a mono-morphemic analysis. Let's call this *the null 3SG account*. We will first explore this latter option and then the first one. Both analyses, we argue, can be defended by reference to properties that play a role in Romanian at large. Although it is not entirely obvious to us what the most viable analysis is, we speculate that the deletion account seems to pose fewer obstacles for language acquisition.

4.2 *The null 3SG account*

Let us look at the null 3SG account in detail. Following Illiescu & Mourin (1991), we could take 3SG *-ă* and *-e* in the present tense to be forms that denote the present indicative rather than person and number agreement. The core argument for this analysis is that in the subjunctive paradigm the 3SG is always marked by an opposite vowel. In the first conjugation, 3SG present tense is *-ă* and the 3SG subjunctive *-e*, whereas in the other conjugations the 3SG present tense form is *-e* and we find *-ă* in the 3SG subjunctive. These contrasts would then justify the feature [indicative] for the 3SG markers in the present tense (or [Mood:] in a system that takes the indicative to be unmarked). It stands to reason to assume that these contrasts are part of the primary input data that children use to acquire Romanian because, in contrast to the other Romance languages, the subjunctive is used as an alternative to infinitival complementation, which makes it very pervasive in the input. The consequence of this analysis is that the 3SG agreement marker in the present tense is null, just like in Italian and Spanish:

(23) $-\emptyset$ $\langle \rangle$ [Agr:]

We can subsequently assume that this null form is also generated in the past tense, as there is no evidence against it. The result of this is that there is no reason to switch to a mono-morphemic analysis because there is no agreement alternation to begin with, let alone one that requires postulation of a null form or homonym, as was the case in English and Icelandic.

In order to make this analysis work, we need to address a couple of issues. First of all, if *-ă* and *-e* are not agreement markers but markers of the present indicative, the question arises why we only see them occur in 3rd person contexts in the first conjugation, and only in 3SG context in the fourth conjugation. This question is especially pressing for the first person, where no overt marker follows the verbal root, as can be observed in (16). Second, if agreement is marked by a null form in the 3SG present tense context, it shares this property with the 1SG present tense context (cf. the analyses in (18) and (22)). This is awkward in light of the fact that 1SG and 3SG contexts can be straightforwardly analyzed as a natural class: they are the singular contexts not marked by [addressee]. An analysis that nevertheless postulates two distinct null forms for these contexts therefore postulates an avoidable homonym pair, and we have argued earlier that children avoid postulating such homonyms. On the other hand, if we treat the null form in the 3SG context as the same one that occurs in the 1SG context (in other words, we treat the null form as an elsewhere form that is also generated in the 1SG context by

lack of an alternative), we beg the question why 1SG contexts do not show the hypothesized present indicative markers *-ă* or *-e*.

These problems can be addressed as follows. The reason why we only see the present indicative markers in a 3SG context is because in all other contexts it would be preceded or followed by a root-external vowel. We can therefore postulate the following deletion rule:¹³

$$(24) \quad -V_{\text{INDICATIVE}} \rightarrow \emptyset / \text{..}]_{\text{ROOT}} _ V _ _$$

The triggering vowel can either be a following agreement vowel, as in 2SG *-i*, or a preceding theme vowel, like *-a* in the first conjugation. In the first conjugation, this accounts for the absence of the indicative marker in 2SG contexts (because of the adjacent agreement marker) and 1PL/2PL contexts (because of the adjacent theme vowel) and its presence in 3SG/3PL contexts, where agreement is hypothesized to be null and the theme vowel is absent. The remaining question is then why the indicative marker does not show up in 1SG contexts, where no obvious marker appears following the root. Focusing on the first conjugation, an answer is readily available, namely the presence of an underlying vowel. It can be argued that the 1SG agreement marker is *-u* but this marker only surfaces in specific phonological contexts, namely after a root-final consonant clusters consisting of an obstruent plus a liquid or after a root-final vowel, where it is realized as a /w/-glide (cf. Chitoran 2002). This is not something that has to be stipulated for 1SG verbs. The exact same contexts account for the realization of *-u* in a subset of masculine/neuter nouns and adjectives. So we have the following minimal pairs for verbs, nouns and adjectives, respectively.

(25)	a.	cânt(-u)	afl-u	bea-u
		sing.PRES.1SG	learn.PRES.1SG	drink.PRES.1SG
	b.	lup(-u)	teatr-u	pusti-u
		wolf.M.SG	theatre.N.SG	desert.N.SG
	c.	frumos(-u)	albastr-u	auri-u
		beautiful.M/N.SG	blue.M/N.SG	golden.M/N.SG

In e.g. the nominal domain, the underlying *-u* comes at the surface in definite contexts, where we see *lupul* (as well as *teatrul*). The *-u* following the root *lup-* cannot be straightforwardly analyzed as an epenthetic vowel, as it would be the only epenthetic *-u* vowel in the grammar of Romanian, Chitoran observes. It should be noted that in the verbal domain the underlying vowel surfaces as *-i* if the stem ends in *-i*, so we find *afl-u* learn.1SG vs. *speri-i* scare.1SG, where *-i* is an allomorph of *-u* and therefore still provides evidence for an underlying vowel.

In short, there are at least three contexts in Romanian where an underlying *-u* can be assumed. However, this *-u* is deleted most of the time and only surfaces in two specific phonological environments (after obstruent+liquid clusters and after vowels). This pattern is the rule in Romanian: it is regular and productive. With the presence of an underlying agreement form *-u* in 1SG contexts, the absence of the indicative marker in these contexts is now accounted for: it is simply another consequence of the deletion rule in (24), with underlying *-u* triggering deletion of the indicative vowel.

Let us see what the implications are for the other conjugations. Here, we again see that the 1SG context does not have an obvious marker following the root (cf. (16) for the fourth conjugation). In contrast to what we see in the first conjugation, the 3PL context and the 1SG are always similar. In order to understand the absence of the indicative marker in 1SG and 3PL contexts alike, we must therefore hypothesize that an underlying vowel must be present in both. And indeed, verbs with an *-u* surfacing in exactly the 1SG and 3PL contexts can be found:

(26) Third and fourth conjugation verbs with 1SG/3PL *-u*

	THIRD CONJUGATION (-E VERBS)		FOURTH CONJUGATION (I-VERBS)
	a scrie ('to write')	a umple ('to fill')	a ști ('to know')
1SG	scriu	umplu	știu
2SG	scrii	umpli	știi
3SG	scrie	umple	știe
1PL	scriem	umplem	știm
2PL	scrieți	umpleți	știți
3PL	scriu	umplu	știu

One might wonder how robust the evidence is for an underlying *-u*. The general issue is that not all verbs with a root ending in a vowel or [obstruent+liquid] consonant cluster take an *-u* ending in 1SG and 3PL contexts. Many verbs take a root extension instead, *-ez* in the first conjugation and *-esc* in the fourth conjugation. These root extensions end with a phonological environment that suppresses rather than promotes the appearance of *-u*. Therefore, evidence for an underlying *-u* marker can only come from verbs that do not take a root extension. To get a rough impression of how robust the relevant evidence is, we used www.cooljugator.com/ro/list, which provides the conjugations of 1292 verbs, including many frequent ones. Table 1 gives an overview of the percentages of verbs providing evidence for an underlying vowel (i.e. verbs not taking root extensions).

(27) Table 1: Evidence for underlying vowel in 1st versus 2nd/3rd/4th conjugations

ROOT ENDING IN:	CONJUGATION 1	CONJUGATIONS 2, 3 AND 4
vowel	59% (n=17)	12% (n=8) ¹⁴
[obstruent+liquid]	68% (n=28)	23% (n=30)
Total	64% (n=45)	21% (n=38)

As you can see, the evidence for an underlying vowel is significantly less robust in conjugations 2, 3 and 4 than in conjugation 1. What is particularly relevant for conjugations 2, 3 and 4 is that the child will have to actively construe a second context in which such an underlying *-u* vowel will have to be postulated, namely the 3PL context. Since the 1SG and 3PL contexts do not form a natural class, the *-u* marking the 3PL context constitutes an entry distinct from the *-u* entry marking the 1SG context. Either the child postulates the same form for the 3PL context by way of analogy (i.e., through the fact that the 1SG and 3PL always display the same form), or the data in (27) suffice for postulation of a 3PL *-u* form. Note what is at stake here. The issue is not whether there is enough evidence for the child to eventually acquire an underlying vowel in 1SG and 3PL contexts successfully. There is. The issue is whether the appearance of the underlying vowel is robust enough at the relevant stage for the child to seriously entertain the possibility of the deletion rule in (24), with indicative markers *-ă* (1st conjugation) and *-e* being blocked by underlying vowels in 1SG and 3PL contexts. In other words, the underlying vowel is just one prerequisite, but an essential one, for the null 3SG account to work.

It is not the only prerequisite. Another issue concerns the distribution of the theme vowel. If *-ă* and *-e* are indicative markers, the child must assume that in 1SG/2SG contexts generating the agreement marker takes precedence over generating the theme vowel, because the theme vowel does not surface there. At the same time, it must assume that in e.g. 2PL contexts generating the theme vowel takes precedence over generating the present indicative marker, as the latter does not surface there.¹⁵ This is especially clear in the fourth conjugation, where in 1/2PL contexts theme vowel *-i* beats indicative marker *-e*. Not only must the child

conclude that the theme vowel interacts differently with distinct inflectional markers, the outcomes of these interactions are especially puzzling in light of the fact that agreement markers make no semantic contribution whereas tense/mood markers do.

Let us sum up the null 3SG account. Under the assumption that the forms we see in 3SG present tense contexts are present indicative and not agreement markers, we can rescue a bi-morphemic analysis for Romanian. A consequence, however, is that we must assume that the child is able to entertain some auxiliary hypotheses that function as prerequisites for the null 3SG account. Although one might argue that the evidence for all these acquisitional steps is there, one might as well argue that this evidence is not very robust overall, especially given the complexity of the intermediate steps. These indeterminacies, then, characterize the weakness of the null 3SG account.¹⁶

4.2 The deletion account

Let us continue with the deletion account. A noteworthy property of Romanian is that vowel reductions happen quite prominently in its inflectional systems. First of all, we already saw for the first conjugation and the (ad)nominal domain that *-u* only shows up in a restricted number of phonologically determined contexts, which means that it is deleted more often than it is not.

Second, it is also clear that we do not always see that two vowels inserted by regular morpho-phonological rules make it to the surface form. Take a look again at the imperfect paradigms in different conjugations. Under the assumption that the imperfect marker is *-a* throughout, the [ea] form we see in the fourth conjugation can be analyzed as a combination of the theme vowel *-i* and the imperfect marker *-a*, as proposed by Chitoran (2002). This analysis can be extended to the second conjugation, where [ea] can be taken as the phonological outcome of theme vowel *-e* with imperfect marker *-a*. On a par with this analysis for the fourth conjugation, she consequently assumes that in the first conjugation theme vowel *-a* and imperfect marker *-a* merge into [a]. The question then becomes what happens when two adjacent vowels, expressing thematic or inflectional properties, are not identical and cannot be diphthongized into an existing diphthong. What happens, for instance, if the vowels involved are *-a* and *-ă*, where neither [aă] nor [ăa] constitutes a possible diphthong in Romanian? The nominal domain provides a clear answer: the outcome of such a merger is [a]. When a feminine noun like *casă* is inflected with a definite marker *-a*, the outcome is *casa*. In contrast, inflecting feminine *lume* with *-a* results in *lumea*, because [ea] is an existing diphthong.

These observations now have non-trivial consequences for the analysis of the agreement system in Romanian. If the 3SG form is *-ă* is transparently attached to the right of the imperfect marker *-a*, it simply cannot survive in that position, as it cannot be diphthongized with imperfect marker *-a*: [aă] is not a possible diphthong. One of the two vowels therefore has to give in and, like is the case in the nominal domain, the [a] vowel wins out. The fact that we do not see the *-ă* agreement marker surface in the imperfect tense is not because it is never inserted but because it has undergone vowel merger with the dominant *-a*. The fact that *-ă* does surface in the 3SG of the present tense must then be due to the absence of the theme vowel in that context: There is no adjacent vowel that *-ă* has to merge with. The absence of the theme vowel in 3SG context does not have to be stipulated, as we do not see it occur in the other singular contexts either. In addition, Chitoran (2002) argues that theme vowels always attract the stress, even after they assimilate with the imperfect vowel *-a*. The fact that 3SG *-ă* does not attract the stress in the present tense therefore counts as a strong indication for the absence of the theme vowel in that context.

For the fourth conjugation, we can essentially extend the analysis we offered for the first conjugation. Under the assumption that 3SG *-e* is regularly attached to the imperfect marker *-a*, the result cannot be diphthongized into *[ae]. What makes matters more

complicated, of course, is that the imperfect marker *-a* is in turn merged with the theme vowel *-i* into the diphthong [ea], so that adding *-e* would in fact create a triphthong. Even if one has a lenient view on the number of triphthongs in Romanian (see the list in Stan 2012:12), [eae] is not on that list.¹⁷ We can therefore assume, as we did before, that in the merger of these vowels [(e)a] is dominant.

To sum up the deletion account, there is no reason to assume that the present tense 3SG *-ă* and *-e* forms are not regularly inserted in the imperfect, despite the fact that we do not see them surface. General phonological processes we can independently observe in Romanian inflectional systems account for it. In contrast to the null 3SG account, the deletion account does not seem to require auxiliary assumptions. Vowel reduction is robust and regular enough in Romanian to do the job for us here. Aware of these regularities in Romanian, the child will not be tempted into thinking that the language (s)he is learning is mono-morphemic. Note that if we want to rescue Icelandic or English from becoming mono-morphemic in a similar way, we would have to show that independent evidence exists in Icelandic for *-r* deletion, and for *-s* deletion in English, in similar contexts. We do not think this evidence exists.

5 Underpecification and intransparency in Romanian

In this section, we will look at two general issues. Both involve consequences of our analysis for pro drop in Romanian. The first issue is the following. Romanian is a rich agreement language but does not always uniquely mark different contexts, in contrast to what Italian and Spanish do in the present tense. The question is whether our proposal now predicts that in contexts without a unique marker pro drop is prohibited, to which our answer will be negative. A second issue has to do with the fact that Romanian is not transparent in its encoding of tense and agreement all the way down. We will show that neither is this the case in Italian and Spanish and argue that this observation does not harm the pro drop possibilities either.

With the bi-morphemic analysis for Romanian in place, the language meets the suggested prerequisite for licensing null subject. Although agreement in Romanian is clearly rich (it has enough contrasts to justify the features [speaker], [addressee] and [plural]), a question arises if it is not problematic that in both conjugations one form can show up in more than one slot, both in the present tense and the imperfect. After all, there are no forms uniquely identifying the unexpressed subject in those cases. It is therefore important to state that pro drop licensing must be syntactic and not phonological: it is in the syntax that in a particular context the feature set expressed on the verb must match the one of the null subject. Now, in the present tense of the first conjugation, the form appearing in the 3SG context is identical to the one occurring in the 3PL context, namely *-ă*. It is not the case, however, that these contexts always display the same form. In the imperfect tense, the form *-u* is inserted in 3PL contexts with the help of a context-sensitive rule (see (18) and (22)). This rule must be able to read the features present on the underlying morpheme if it wants to successfully insert *-u*. It must therefore be the case that [plural] is part of the morpheme that is generated in all 3PL contexts, despite the fact that it is not spelled out by a specific form in the present tense. This means that even in the present tense a 3PL null subject is licensed by the presence of this plural feature. The consequence is that the sentence *Cântă!* is ambiguous between '(S)he sings!' and 'They sing!' but it is not ungrammatical. The same is true for the fact that the form *-m* occurs in both 1SG and 1PL imperfect contexts. The form *-m* spreads to the 1SG context in the imperfect due to impoverishment of the [singular] feature in the imperfect, but this feature can only be impoverished if it is there in the first place. This means that the syntactic representation must include a morpheme in 1SG imperfect contexts that includes a [singular] feature, which is distinct from the morpheme used in 1PL contexts, which includes [speaker] but not [singular].

This is enough to license a 1SG or 1PL null subject. The fact that [singular] is subsequently impoverished cannot undo this. As a consequence, the sentence *Cântam!* is ambiguous between *I sang!* and *We sang!* but it is not ungrammatical.

Let us now turn to the second issue. The fact that for Romanian the learner can come to a bi-morphemic analysis of tense and agreement and finds no urgent reason to switch to a mono-morphemic representation does not necessarily entail that Romanian encodes tense and agreement transparently all the way down. As is the case for the Italian ‘passato remoto’ and the Spanish ‘pretérito perfecto’, the simple perfect paradigm offers a less transparent picture, primarily due to the fact that there is no obvious marker for the perfect, distinguishable from the theme vowel.

(28) Imperfect and simple perfect paradigms in Italian, Spanish and Romanian

	ITALIAN		SPANISH		ROMANIAN	
	imperfect	simple perfect	imperfect	simple perfect	imperfect	simple perfect
1SG	amavo	amai	amo	amé	cântam	cântai
2SG	amavi	amasti	amas	amaste	cântai	cântași
3SG	amava	amò	ama	amó	cânta	cântă
1PL	amavamo	amammo	amamos	amamos	cântam	cântarăm
2PL	amavate	amaste	amáis	amasteis	cântați	cântarăți
3PL	amavano	amarono	aman	amaron	cântau	cântară

One could of course argue that in these paradigms the perfect marker is null but that immediately creates a tension. Under the assumption that null forms in general are paradigmatically licensed, it does not help that the present tense is already expressed by a null form. Alternatively, one could argue that the tense and agreement morphemes undergo fusion, so that one form, expressing tense and agreement features can be inserted. Arregi (2000) postulates such an analysis for Spanish with the following two arguments: (i) the inflectional material following the stem (the root plus theme vowel) is more condensed in the simple perfect and (ii) we find agreement forms that are specific to this tense. With respect to (i), note for instance the difference between *am-a-b-a* (1/3SG imperfect) and *am-é* (1SG simple perfect). With respect to (ii), the *-é* in the 1SG simple perfect cannot be broken down into a tense and agreement form in any transparent way, so it seems to function as both a tense and agreement marker at the same time: it is a portmanteau form. In Romanian, (i) is not obviously the case. Note that in the imperfect Spanish has two theme vowels in the first conjugation, one preceding and one following the consonant marking the imperfect tense (*cantabas* = *cant*_{STEM}-*a*_{THEME}-*b*_{IMPERFECT}-*a*_{THEME}-*s*_{2SG}). In the Romanian simple perfect, however, the singular forms look similar in complexity to those in the imperfect, whereas the plural forms in fact look more complex in the simple perfect. However, (ii) appears to have examples in Romanian too. The inflectional material following the theme vowel *-a* in 1SG and 2SG contexts are *-i* and *-și* and we do not find these forms in the same contexts in either the present or imperfect tense. Under the assumption that in the simple perfect T and Agr also fuse in Romanian, we need spell-out rules that refer to both tense and agreement features, like the following:

- (29) *-i* <> [[Agr: speaker], [T: past^{PERFECT}]]_{T+Agr}
-și <> [[Agr: addressee], [T: past^{PERFECT}]]_{T+Agr}

Suppose that we settle for such a fusion analysis. Does that impact the possibility of licensing null subjects? The answer is negative. It is important to realize that fusing tense and

agreement is distinct from a mono-morphemic analysis. After all, T and Agr would not be able to fuse if they did not exist in the first place. The crucial difference between Italian, Spanish and Romanian on the one hand and Icelandic, German and English on the other, then, is that the former provide unambiguous evidence for a morphemic separation of tense and agreement (namely, the contrast between the present and imperfect tense), whereas that evidence is lacking in the Germanic languages. The fact that the Romance pro drop languages are not transparently bi-morphemic throughout all paradigms is therefore immaterial to the predictions.¹⁸ The requirement is not that every paradigm should provide evidence for the syntactic representation in (5)a, the requirement is that non-peripheral data in the language should.

6 Conclusion

Given previous literature on the pro drop languages, the fact that Romanian is one can hardly be considered surprising. The pervasive generalization about agreement-based pro drop has for a long time been that agreement has to be rich. When one looks at the morpho-phonological contrasts in the Romanian agreement paradigm and at the spell-out rules of the grammar that would correctly generate these forms (cf. (18) and (22)), then it is clear that these contrasts ensure the presence of the agreement values [speaker], [addressee] and [plural] in the Romanian agreement system, which can then be combined to provide fully specified agreement morphemes. These morphemes consequently allow the licensing of null subjects. Even an impressionistic analysis could conclude that. The issue, however, is not how to understand Romanian with respect to pro drop, but how to understand this language in contrast to e.g. Icelandic. This language shares with Romanian that it has five morpho-phonological contrasts in the present tense and that the only two slots with similar forms become distinct in the past tense. If Romanian has pro drop, so should Icelandic, contrary to fact. This puzzle must be addressed if pro drop theory is to have any predictive power. We have proposed a new way of looking at pro drop and have worked out the consequences for Icelandic and in particular Romanian. Our investigation highlights two general points about this type of research.

First of all, counting contrasts impressionistically is insufficient for making the right predictions. Having particular, and enough, morpho-phonological contrast might be a prerequisite for pro drop, but it does not guarantee the licensing of null subjects. It is only after a fine-grained analysis of particular languages that specific properties are noticed that lead to new hypotheses, such as the relevance of the bi-morphemic expression of tense and agreement.

Second, even in the light of a specific hypothesis about an additional prerequisite, a casual look at a particular language to see if this prerequisite is met does not suffice either. We argued that the impossibility of licensing null subjects in Icelandic and English was due to the fact that these languages have a conflated category for tense and agreement, INFL, which is consequently overspecified for the purposes of licensing an empty subject. The trigger for such an analysis in these languages is the fact that the 3SG present tense form (*-ir* in Icelandic, *-s* in English) does not come back in the past tense. This is in contrast to Italian and Spanish, where the 3SG context is marked by a null form in both the present and imperfect tense. Now interestingly, we saw for Romanian that the 3SG form in the present does not reappear in the 3SG imperfect context either, so that it actually seems to pattern with Icelandic and English. It was only after a careful look at inflectional morphology in Romanian that we could establish that phonological reduction processes are very pervasive in the inflectional domains of the language. The disappearance of 3SG *-ă* in the imperfect tense can therefore be taken as another example of just that.

In order to understand the distribution of null subjects in the world's languages, one must understand the general patterns across languages as well as the fine-grained differences between minimal pairs. In this paper, we have sacrificed the first in order to make a contribution to the second enterprise. This exercise, we think, is important, as it leads to more precise hypotheses and it ensures that in our hypothesis-testing on a larger scale we do not take a particular language as a straightforward falsification too hastily.

References

- Aalberse, Suzanne & Jan Don. 2009. "Syncretism in Dutch Dialects". *Morphology* 19:1, p. 3-12.
- Allen, Andrew. 1977. "The interfix I/ESC in Catalan and Rumanian". *Romance Philology* 31, 2: 203–11.
- Arregi, Karlos. 2000. "How the Spanish verb works". Handout of a talk given at LSRL 30, at the University of Florida, Gainesville.
- Bayer, Joseph. 1984. "COMP in Bavarian syntax". *The Linguistic Review* 3: 209–274.
- Benveniste, Émile. 1971. *Problems in general linguistics*. Coral Gables, FL: University of Miami Press.
- Bobaljik, Jonathan & Höskuldur Thrainsson. 1998. "Two heads aren't always better than one". *Syntax* 1: 37-71.
- Bonet, Eulalia. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Cambridge, MA: MIT dissertation.
- Calabrese, Andrea. 2015. "Irregular Morphology and Athematic Verbs in Italo-Romance". *Isogloss. A Journal on Variation of Romance and Iberian languages*. 69-102.
- Calabrese, Andrea. 2019. Morpho-Phonological Investigations: A Theory of PF. From Syntax to Phonology in Sanskrit and Italian Verbal Systems. Ms., University of Connecticut.
- Chitoran, Ioana. 2002. *The Phonology of Romanian: A Constraint-Based Approach*. Berlin/New York: Mouton de Gruyter.
- de Haan, Ger. 1994. "Inflection and cliticization in Frisian". *NOWELE* 23: 75–90.
- Forchheimer Paul. 1953 *The category of person in language*. Berlin: Walter de Gruyter.
- Harley, Heidi. & Elizabeth Ritter. 2002. "A feature-geometric analysis of person and number". *Language* 78: 482-526.
- Iliescu, Maria and Louis Mourin. 1991. *Typologie de la morphologie verbale romane*. Innsbruck: Amoe.
- Koeneman, Olaf and Hedde Zeijlstra. Under review. Pro drop and the morphological structure of inflection. Ms. Radboud University/ Georg-August-Universität Göttingen.
- Oltra-Massuet, Isabel. 1999. "On the Constituent Structure of Catalan Verbs". In K. Arregi, B. Bruening, C. Krause, and V. Lin, eds., *Papers in Morphology and Syntax, Cycle One*, volume 33 of MIT Working papers in Linguistics, MITWPL, Cambridge, MA, 279–322.
- Preminger, Omer. 2014. *Agreement and its failures*. Cambridge, MA: The MIT Press.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Saab, Andrés. 2008. *Hacia una teoría de la identidad parcial en la elipsis*. Doctoral dissertation, University of Buenos Aires.
- Stan, Camelia. 2012. "Phonological and orthographic features of Romanian". In Dindelegan, G. P. (ed.) *The grammar of Romanian*. Oxford: Oxford University Press.
- Taraldsen, Knut Tarald. 1978. *On the NIC, vacuous application and the That-Trace Filter*. Ms., MIT.
- Zafiu, Rodica. 2012. "Mood, tense, and aspect". In Dindelegan, G. P. (ed.) *The grammar of Romanian*. Oxford: Oxford University Press.

¹ This means that we believe that radical pro drop, as observed in languages like Chinese and Korean, must be treated as a different phenomenon, and indeed it has different characteristics. See Neeleman & Szendrői (2007) for discussion.

² For some poorly understood reason, an agreement marker can license an empty subject even if the corresponding subject is marked for gender but the agreement marker is not, at least in the Romance languages. We will ignore this issue.

³ In fact, this property is shared by all Germanic languages with agreement and these languages crucially lack null subjects (modulo the partial pro drop languages mentioned in the previous section). In Koenen & Zeijlstra (under review) we argue that partial pro drop in Frisian and Bavarian dialects is licensed by an additional property absent from the standard languages, namely complementizer agreement. It is generally the case that partial pro drop only occurs in contexts that in embedded clauses would show complementizer agreement. We argue that the presence of such an agreement form promotes an analysis with a separate agreement morpheme in that context, which can then in turn license a null subject. See the reference for details.

⁴ See Oltra-Massuet (1999), Arregi (2000) and Calabrese (2015) for details.

⁵ Interestingly, Standard French has a *sjwa* in the 3SG present tense (*demande* /dəmãd(ə)/), whereas there is no evidence of this *sjwa* returning in the 3SG past context, which has *demandait* /dəmãdɛ/. This predicts that Standard French should pattern like English and Icelandic, and indeed it has no argumental pro drop. See Koenen & Zeijlstra (under review) for the details of this analysis.

⁶ An alternative analysis is offered by Calabrese (2019), who takes *-ya* to be a theme vowel allomorph of *-a* and the 1PL agreement marker to be *-mo*. In this way, he is able to transparently maintain the presence of a theme vowel throughout the plural part of the paradigm.

⁷ Note that both the bi-morphemic and mono-morphemic analyses of English need an additional null form to account for the plural ($-\emptyset \langle \text{Agr: plural} \rangle$ and $-\emptyset \langle \text{I: present, plural} \rangle$, respectively). The overall difference between the two analyses is therefore four versus two null forms, favouring the mono-morphemic one.

⁸ Since one of the contexts in which *-t* appears is 3SG, one could of course analyze both *-t* forms as instantiations of the same elsewhere form. Observe, however, that in the past tense this *-t* disappears in the 3SG context (which has *spielte*) but not in the 2PL context (which has *spillett*). This constitutes another argument for treating the two *-t* forms as accidentally homonymous.

⁹ We will only consider the so-called strong forms and not the weak forms that have semantically empty stem extensions *-ez* (1st conjugation) and *-esc* (4th conjugation) in all singular and 3PL contexts (cf. Allen 1977 and Zafiu 2012:26 for further discussion). They do not impact our overall assessment, as far as we can tell.

¹⁰ In the imperfect tense, both *-m* and *-u* are now possible spell outs of the agreement morpheme [Agr: speaker, plural]. To ensure that *-m* wins, we need to assume that spelling out a person feature is more specific than spelling out a number feature, which makes sense in a system in which number has two dimensions and person has three (i.e., person is more restrictive).

¹¹ The analysis in (18) overgenerates in that it expects the theme vowel to show up in any present tense slot, whereas in fact it is restricted to 1PL and 2PL contexts. More needs to be said about the distribution of theme vowels, which is poorly understood. Note for instance that one could try to argue that the theme vowel does not show up in singular and 3PL contexts because all agreement endings following are vowels. However, there are reasons to believe (see below) that in the imperfect tense the theme vowel is inserted to the left of the imperfect marker, which is also a vowel. We will refrain from giving a complete analysis, also because our analysis does not require one.

¹² One technical issue needs solving. We have assumed throughout that singular denotes the absence of plural, much like 3rd person denotes the absence of person. At the same time, we have postulated the feature [singular] as part of the analysis of the 1SG form in Romanian (see (18)). This allowed us to impoverish this feature in the imperfect, with the consequence that another form, *-m*, takes its place. Including [singular] as a feature must therefore constitute a marked analysis. The issue is now as follows: If the Romanian agreement patterns allow us to postulate a singular feature, why would we not postulate it for the 3SG contexts in Icelandic and English? We can then subsequently impoverish this feature in the past tense, ensuring that *-ir* and *-s* can no longer be inserted there. This would allow one to maintain a bi-morphemic analysis for these languages and pro drop is subsequently expected, an unwanted result. Note, however, that if we block the spell-out rules referring to *-ir* and *-s* from

applying, no morpho-phonological form can be inserted at all, leading to a crash. This crash can only be avoided by postulating a null allomorph for these 3SG past tense contexts, and we are back at square one. In short, including the feature [singular] in the analyses for Icelandic and English would not solve the problem but merely reinstate it.

¹³ The hypothesized indicative marker is not deleted in the context of a root-final vowel (cf. (26), where we see that the 3SG form for *a ști* is *știe*), hence the inclusion of `..]ROOT_` in the context description. We follow Chitoran (2001) in assuming that theme vowels are not part of the root.

¹⁴ We counted *prescrie*, *rescrie*, *scrie*, *subscrie* and *transcrie* as one example.

¹⁵ Assuming that the theme vowel *-a* is inserted in 3SG contexts but assimilates with the present indicative marker *-ă* into *-â* is suspect, as there are strong reasons to assume that, in the combination of *-a* and *-ă*, *-a* is in fact the dominant vowel. See section 4.2 for discussion.

¹⁶ A secondary problem for the null 3SG account created by the conjugations other than the first one is that the verbs that the verbs in (26) do not show a contrast between the 3SG present indicative and the 3SG subjunctive, where *-e* is used across the board. This will of course not help the child to establish that *-e* is the indicative marker, but overall this pattern is the exception.

¹⁷ There is discussion about what counts as a real diphthong or triphthong in Romanian. Chitoran (2001:8-9) argues that only /*ɛa*/ and /*ɔa*/ count as diphthongs and that additional cases should be analyzed as vowel-glide combinations. One's specific view on this does not affect our argument, as far as we can tell.

¹⁸ What reduces the relevance of the simple perfect in Romanian further is that, apart for some regions, this tense is generally replaced by a periphrastic one (cf. Chitoran 2001:62).