How weak are Romanian clitic pronouns?

Ciprian-Virgil Gerstenberger
Norwegian Centre for E-health Research

Abstract

In traditional linguistics, pronouns are divided into two classes: those that can bear word stress, coined "strong", "full" or "tonal", and those that cannot, coined "weak", "clitic", or "atonal". However, in the last decades, research on this topic has shown that items generally labeled as clitics are far more complex. Somewhere between words and affixes, these hybrid linguistic entities challenge both description and modeling. As for Romanian, the debate on weak (i.e., clitic) pronouns has been dominated by the question of their categorial status: are these items clitics or affixes? In this article, I present and scrutinize different approaches that support the claim that there are differences between proclitics and enclitics, i.e., between clitics occurring before vs. after the verb; this includes not only positional, but also featural differences. I identify various types of ambiguities in Romanian that could lead to improper data interpretation, and, based on an analysis of syllabicity – the most salient feature of Romanian weak pronouns – I refute claims for treating clitics in preverbal position differently than in postverbal position. Furthermore, using evidence from both historical data and data pertaining to language varieties, I show regularities in the Romanian weak pronoun system, bringing evidence against the claim that Romanian weak pronouns show a great deal of idiosyncrasies.

Keywords: Romanian, clitics, clitic pronouns, weak pronouns, description, rule-based model

1. Introduction

Browsing through the vast literature on clitics, one can hardly find an article or a book that does not highlight the complexity of the topic and the difficulties of attempting to pin down the very nature of clitics. Somewhere between words and affixes, the label "clitic" has been applied to items that are not clearly words, nor clearly affixes (cf., for instance, Spencer and Luís 2012).

The originally simple concept of an item 'leaning' on a host either from one side, as a proclitic, or the other, as an enclitic, was not enough to describe the extensive variety of hybrid forms and their occurrences. Even the fact that clitics may occur in a clitic cluster challenges the idea of clitic-hood. In a sequence of two clitics preceding the host, it is only the second clitic which truly 'leans' on the host, while the first clitic can only 'lean' on the following clitic. The same issues with the simple clitic-host notion arise around what is referred to as mesoclitics in Portuguese (cf. Rouveret 1999), or endoclitics in Udi[1] (cf. Harris 2002). Conceptually, one and the same clitic cannot 'lean' on both sides of a host at the same time, but this is the case both with mesoclitics – which occur between verb stem and affixes – and with endoclitics – which split the root into two parts: both constructions challenge the Lexical Integrity Hypothesis (cf. Di Sciullo and Williams 1987) unless they are regarded as affixes – as Haspelmath (2022:p. 31) does with the items in the verbal complex in the Romance languages. As I will show in Section 2, for Romanian, these problems – intensified by different types of ambiguities in the orthographic system – have led to difficulties in accurately describing the clitic-host relationship.

1.1. Description issues

There seems to be a general pattern in detailed descriptions of clitics, according to which the assertion of a clitic property is accompanied by exceptions, as the following list exemplifies (cf. Caink 2006):

1. **Rigid order** Clitics, as affixes, appear in rigid order, yet there are counterexamples, e.g., Bonet (1991) for Catalan or Săvescu Ciucivara (2009) for Romanian.
2. **Stress** One of the most prominent features of clitics and the feature that coined the terms 'weak' or 'atonal' for pronouns is stress. Clitics cannot bear accent or stress, though there are plenty of counter-examples of stressed clitics in the vast literature, e.g., Klavans (1995), Ordóñez and Repetti (2006) or Săvescu Ciucivara (2009).

[1] Udi is a Nakh-Daghestanian language of Azerbaijan and Russia.

©2022 Ciprian-Virgil Gerstenberger. *Nordlyd* 46.1: 37-57, *Morfologi, målstrev og maskinar: Trond Trosterud {fyller | täyttää | deavdá | turns} 60!*, edited by Lene Antonsen, Sjur Nørstebø Moshagen and Øystein A. Vangsnes. Published at UiT The Arctic University of Norway. http://septentrio.uit.no/index.php/nordlyd https://doi.org/10.7557/12.6341

This work is licensed under a Creative Commons "Attribution-NonCommercial 4.0 International" license.

3. **Coordination** Syntactic rules of coordination never apply for clitics, and yet, examples of coordinated clitics are found in the literature, e.g., Finocchiaro (2005) for French or Săvescu Ciucivara (2009) for Romanian.
4. **Ellipsis** Clitics are not affected by ellipsis, yet there is evidence for the opposite, e.g., Franks and King (2000) for Serbo-Croatian or Finocchiaro (2005) for Italian.

Based on these descriptions, it is not clear whether the assertion that some clitic behavior contradicts the general view on clitics (e.g., clitic coordination in Romanian) is actually a part of the standard language or it is only the idiosyncratic view of a linguist or of a linguist's informant(s). This circumstance becomes apparent for non-native speakers when discrepancies between native speakers' opinions about the grammaticality of a language sample emerge (see, for instance, Bošković 2001:p. 122, footnote 25).

Cardinaletti and Starke (1999) propose a tripartite model that distinguishes between strong, weak and clitic pronouns in terms of *structural deficiency*. They argue that Italian stressed enclitics are weak pronouns distinct from other clitics. Yet, Manzini (2014) maintains that it is "not obvious that the intermediate series ('weak') displays consistent characteristics" (for a similar view, see also Pescarini 2018). Moreover, Cardinaletti and Starke (1999)'s tripartition of pronouns, deemed by the authors as *universal*, does not fit the Romanian data, as argued for in Somesfalean (2007:p. 6). The description of Romanian data in this study brings additional support to Haspelmath's (2015) view on Cardinaletti and Starke's (1999) model of structural deficiency: "their tripartition of person forms into *clitics*, *weak pronouns* and *strong pronouns* is really based on a few interesting converging observations for German and Italian, and cannot be extended to many other languages without encountering the familiar problems" (op. cit. p. 276, footnote 2).

Regarding clitic coordination, Săvescu Ciucivara (2009:p. 24) argues that, in Romanian, "certain pre-verbal clitics can be coordinated (though not all speakers accept it)". A different perspective is offered by Dindelegan (2013:p. 388), who states that "[i]n informal registers, Romanian […] allows auxiliary and second clitic deletion in coordinated structures. […] gapping of the verb is allowed if it is repeated in the two coordinated structures." Again, there is a report on some speakers' judgements[2] on clitic coordination in Cardinaletti (1999:p. 39): "Benincà and Cinque [(1993)] report that under special prosodic and pragmatic conditions, some speakers of French and Rumanian accept coordination of two clitic pronouns in proclisis contexts, but never in enclisis contexts." Without mentioning any special prosodic or pragmatic conditions, Luraghi (2017:p. 190) claims that "proclitics can be coordinated in French and Romanian." Interestingly, while for Dobrovie-Sorin (1994:p. 61), the coordination of preverbal clitics is "marginally acceptable", for Avram (1997:p. 159), it is "not advisable".[3] In turn, Dobrovie-Sorin and Giurgea (2013:p. 262) asserts that "[c]litics cannot be coordinated, modified or focalized". It is also worth mentioning that the vast majority of Romanian grammars for native speakers do not mention this phenomenon at all. Since there are – to the best of my knowledge – no corpus- or usage-based studies on clitic coordination in Romanian, it is hard to get a clear picture.

Concerning stressed Romanian clitics, Săvescu Ciucivara (2009:p. 24) states that "[w]hen coordination does happen, both clitics are stressed", a statement that contradicts the claim that "[c]litics do not have a word accent" made by Dobrovie-Sorin and Giurgea (2013:p. 262).

As for rigid order, Săvescu Ciucivara (2009) claims that in postverbal contexts, sequences of case-syncretic plural clitic forms *ne*$_{\text{1.pl.acc/dat}}$ and *vă*$_{\text{2.pl.acc/dat}}$ – both *GERUND-ne-vă* and *GERUND-vă-ne* – are grammatically accepted, however, only with the interpretation Acc>Dat,[4] a pattern that contradicts the Dat>Acc positioning of Romanian clitics. This statement has also been mentioned by Dobrovie-Sorin and Giurgea (2013:p. 260). Yet, this allegedly grammatically correct structure is contradicted by any grammar of Romanian, and, as a matter of fact, also by Dobrovie-Sorin and Giurgea (2013) themselves in a book section titled "The order of co-occurring dative and accusative clitics": "Regardless of their position relative to the verb (pre- or post-posed), clitic pronouns form a syntactic unit inside which the order of clitics is fixed, and no insertion is allowed" (op. cit. p. 256). Since the grammaticality judgments of these examples "are based on elicitation and not on corpus research" (Dobrovie-Sorin and Giurgea 2013:p. 260),[5] a series of questions arises, questions that also apply to the claims on Romanian clitic coordination as reported in Cardinaletti (1999:p. 39). Was it an online survey or an experiment? What was the exact number of participants? What was their background and education? If it was a psycholinguistics experiment, was the presentation of examples timed or self-paced? Which distractors were used to prevent participants from giving biased answer, and how were these used? What do statistical analyses result in? Compared to Săvescu Ciucivara (2009), Nevins and Săvescu (2008) provide a clear description of

[2] as reported in Benincà and Cinque (1993)

[3] "[C]oordinarea a două forme neaccentuate […] **nu este recomandabilă**" (my emphasis).

[4] Throughout this study, I use ">" to mark the relative position in the string between two items such as Dat>Acc: the dative clitic occurs before the accusative clitic.

[5] while Săvescu Ciucivara (2009) does not describe the experimental setup of the data collection nor an evaluation of the outcome figures

the experimental setup, outcome figures, and evaluation. It is even more remarkable that, in Nevins and Săvescu (2008), there is absolutely no mention of the acceptability testing of combinations of *ne* and *vă*.

1.2. Terminology issues

In descriptions of the phenomena under scrutiny, there are other difficulties, that related to the use of terminology. For instance, van Riemsdijk (1999a:p.20) labels specific pronouns in Dutch and other Germanic languages as "hostless clitics". Yet conceptually, a clitic requires a host to *lean* on, as this is one of its defining features, and otherwise calling it a clitic does not make any sense. To take another example, synonyms for the same term can suddenly have different meanings due to re-defining them for a specific theoretical model, as is the case with the terms "weak pronoun" vs. "clitic pronoun" in Cardinaletti and Starke (1999). These two terms are used synonymously otherwise, and they are still used as synonyms in the literature that doesn't adopt Cardinaletti and Starke's (1999) view. Without any explicit indications, the reader has to discern between different contexts of use in order to understand the intended meaning of the terms on his/her own. In particular, while for Klein (2007) clitics form a proper subset of weak pronouns – namely only those items that 'phonologically clitizise' – for Cardinaletti and Starke (1999), clitics and weak pronouns are disjointed sets.

In the same vein, Cherecheş (2014) employs terms such "affixal clitics" or "internal clitics" for her prosodic model, despite a footnote in which she mentions "that this usage of the term "clitic" is purely phonological, divorced from the morphosyntactic properties of the element in question" (op. cit. p. 57, footnote 9). Nevertheless, this very specific usage of the term "clitic" obfuscates a clear distinction between syllabic and asyllabic items in the verbal complex, hence veiling a clear distinction between *supporting* and *supported* items. In particular, in the model advocated by Cherecheş (2014) "the auxiliary acts as an affixal (en)clitic when preceded by a pronominal" (op. cit. p. 57), whereas in terms of syllabicity, the vowel-initial auxiliary is always the syllabic host when preceded by a pronominal (e.g., ex. 35, and see the description of obligatory sandhi in Section 2.4.2). Conversely, Dobrovie-Sorin (1994:p. 49) promotes the term "syntactic clitics", while Legendre (2001) the term "verbal clitics" for those entities that Zwicky (1977) labeled "special clitics", e.g., clitic pronouns in Romance.

Furthermore, similar to mesoclitics in Portuguese, Romanian exhibits some structures for expressing wishes, curses, or blessings – remnants of older stages of the language – as referred to in Dobrovie-Sorin and Giurgea (2013:p. 253), Bošković (2001:p. 123, footnote 27), Dindelegan (2016), or Hill and Alboiu (2016). These structures are hardly ever described as mesoclitics in Romanian literature. The reason for this is that, in the configuration VERB>CL-PRON>VERBAL, the VERBAL item is interpreted as an inflected auxiliary verb (the form *ai* in the following Romanian example), and not as an affix of the infinitive main verb (the form *ias* in the Portuguese counter-example), as commonly found in the literature on Portuguese: Romanian *cumpăra-mi-ai cărți* vs. Portuguese *comprar-me-ias livros*[6] 'you would buy me books/may you buy me books'. In international literature however, the very same Romanian structures are inconsistently addressed, either as mesoclitics in Gerlach (2002:p. 57) or as endoclitics in Dobrovie-Sorin (1994:p. 78-79). As already mentioned, mesoclitics occur between verb stem and affixes, while endoclitics split the verb root in two parts. Linguists working with Portuguese or Udi would disagree on the synonymous use of these two terms, which highlights the confusion surrounding analyses of Romanian clitic pronouns.

1.3. Classification issues

Given this description of the various problems and contradictions concerning items labeled *clitic*, it is intriguing to try to bring order into the realm of clitic-hood, i.e., to attempt to identify – in a consistent way – sub-classes of phenomena among items referred to as clitics.

A pioneering account for systematization of the items under the umbrella term *clitics* is Zwicky (1977), who identifies three common types: (1) *special clitics* – unaccented bound forms that are variants of free forms, yet with a 'special' syntax (e.g, Romance clitic pronouns), (2) *simple clitics* – unaccented, phonologically reduced variants of full forms, occurring in the same position as the phonologically full form (e.g., the English *'ll* as in *I'll do it* as opposed to *I will do it*); and (3) *bound words* – always unaccented and phonologically subordinated to a neighboring word (e.g., English genitive *'s*).

In their influential article, Zwicky and Pullum (1983) proposed a suite of tests to distinguish between clitics and affixes (including host selectivity, arbitrary gaps, morphophonological idiosyncrasies, etc.), but Anderson (2005)

[6]example retrieved online 2022-03-10 at URL https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/pronomes-mesocliticos/7865

critically states that "these points are merely descriptive observations about differences in the behavior of two pre-systematically understood classes" (op. cit. p. 33). Despite the fact that these criteria present only tendencies useful for defining what is or is not a clitic, many linguists have used them as solid arguments for discriminating between clitics and affixes (for Romanian, e.g., Monachesi 2001). Even Zwicky himself highlights the distinctions between diagnostic tests and defining criteria in Zwicky (1985:p. 285): "what is normally intended, when such tests are appealed to, is more analogous to medical diagnosis than to operations using an axiomatic system." An overview of different taxonomic efforts towards a general characterization of the category of clitics is given in Haspelmath (2015:p. 275). At the same time, Haspelmath's article is a useful point of reference for an in-depth consideration of the use of grammatical terms cross-linguistically by means of clitic-hood.

In terms of syntactic categories, clitics can belong to various types, such as pronouns, determiners, auxiliaries, negation markers and interrogative markers (cf., e.g., Zwicky 1977). With the complexities concerning clitics and clitic-hood in mind, as presented here and in the previous sections, the remainder of the present study focuses on a single part-of-speech category in a specific language, namely Romanian pronominal clitics, which are weak pronouns in the context they occur in (i.e., the verbal complex).

2. Romanian Weak Pronouns

2.1. Previous approaches

For a long time, the focus of research on Romanian clitics has been on the pronominal items with various labels such as "atonal", "clitic", or "weak", depending on the source. The linguistics literature on Romanian weak pronouns (hereafter RWPs) features a great diversity of descriptions and models. One of the first in-depth corpus-based descriptions of the Romanian verbal complex is provided by Bredemeier (1976), who also works out a detailed and accurate theory-neutral formalization in terms of context-derived constraints. Avram (1986) offers a broad depiction of sandhi in RWPs. Barbu (1999) provides a description of the verbal complex, Dobrovie-Sorin (1999a) a generative approach to the syntax of RWPs, and Somesfalean (2007) an approach to argumental pronominal forms based on data from Romanian and other Romance languages couched in the theoretical framework of the Minimalist Theory. Săvescu Ciucivara (2009) offers another generative perspective on the syntactic analysis of pronominal clitic combinations and ordering in Romance, especially in Romanian. Calude (2001) compares Romanian to French and Serbo-Croatian clitics and concludes that Romanian clitics share many more features with their Serbo-Croatian than with their French counterparts. Furthermore, Monachesi (2001) and Monachesi (2005) deal with RWPs using Head-Driven Phrase Structure Grammar, Barbu and Toivonen (2018) models direct object RWPs within a Lexical Functional Grammar framework, while Klein (2007) treats them within the Dynamic Syntax formalism. Optimality Theory is represented by a series of models such as Popescu (2000), Sasaki and Căluianu (2000), Legendre (2001), Popescu (2003), and Chereches¸ (2014).

There is a long-standing debate about the categorial status of RWPs, specifically concerning whether they are clitics or affixes. While Barbu (1999) and Monachesi (2001) put forward arguments for classifying RWPs as affixes, Popescu (2003) and Gerlach (2002) label them as clitics, irrespective of the RWPs' position towards the verb, i.e., both proclitics and enclitics have the same categorical status as either affixes or clitics. However, some scholars such as Benincà and Cinque (1993) or Cardinaletti and Starke (1999) have promoted the view that there is a fundamental difference between clitic pronouns occurring pre-verbally and those occurring post-verbally; in other words, positing not only a difference in items' position relative to the verb, but also a difference in their other properties.

The threefold aim of the following exploration of the syntax and phonology of RWPs is to present evidence that: (1) there is no basic feature difference between pre- and post-verbal RWPs; (2) it is possible to do away with alleged idiosyncrasies; (3) a thorough, careful description of the data provides a solid foundation for the implementation of a computational linguistics model for generating RWP surface forms.

Note that Appendix 1 and Appedix 2 present an extensive list of examples to provide an overview of the data and facilitate comparisons of individual instances. Unless otherwise specified, example numbers throughout the paper refer to the examples in these appendices.

2.2. Two levels of clitic-hood

The Romanian pronominal system is best described as a tripartite model: strong pronouns (not subject of this study) vs. weak (or clitic) pronouns. The weak pronouns surface in two different forms: syllabic vs. asyllabic (see Table 1).

Due to the ambiguities described below in Section 2.5, the difference between a syllabic and an asyllabic RWP is not always manifest. This issue leads to an inadequate classification of RWPs in terms of syllabicity, which in turn, leads to difficulties in clearly differentiating between host and clitic at the morphophonological level. Moreover, occasionally in the literature on RWPs, it is not clear whether it is about a host at the morphophonological level or at the phrasal level: is the host-clitic pair described in terms of syllabicity or in terms of phrasal stress?

Both Dobrovie-Sorin (1999b) and Klein (2007) point out the need to differentiate between two levels of RWP descriptions that are independent of each other: a phonological/morphophonological level and a syntactic level. The two-level view on ciltic-hood for Romanian presented in this study is congruent with their view.

At the phrasal level, all weak pronouns and other non-stressable – mostly monosyllabic – items in the verbal complex are clitics to the verb,[7] i.e., the prosodic phrase host. This means that, at the morpho-phonological level, an asyllabic RWP form can be a 'clitic' to a neighboring syllabic RWP – the 'syllabic host' – and at the same time, both are proclitics or enclitics to the verb at the phrasal level.

As for the syllabic level, instead of using the terms "phonological" or "morpho-phonological" cliticization, I use the general term "sandhi" and explicitly mark the syllabic support as "syllabic host" and the asyllabic item as "asyllabic clitic" to avoid possible ambiguities.

2.3. Syntactic features of Romanian Weak Pronouns

As with clitic pronouns in other Romance languages, RWPs occur in the verbal complex before the verb as proclitics or after the verb as enclitics. They can form a cluster of up to three RWPs with a fixed order within the clitic cluster: Dat>Dat>Acc, independently of the relative position of the RWP cluster towards the verb (see Dobrovie-Sorin and Giurgea 2013:p. 257ff or Jianu 2013).

In preverbal position, only auxiliaries and a small set of monosyllabic adverbs can intervene between RWPs and the main verb, while in postverbal position, RWPs immediately follow the main verb. In modern Romanian, the auxiliary occurs mostly in preverbal position, however, in wishes, curses, or blessings as well as in vernacular language, the auxiliary may also occur in postverbal position (cf. Gerlach 2002:p. 57). The order of the RWPs cluster and the auxiliary is the same both preverbally (ex. 36) and postverbally (ex. 37), namely RWPs>AUX.

Figure 1 shows an instance of a verbal complex with preverbal RWPs in a maximal context of other possible occurring items, such as negation, auxiliary, monosyllabic adverb intensifiers. Adapted from Chereche̦ş (2014:p. 51) and Dobrovie-Sorin and Giurgea (2013:p. 257), the example shows the fixed order of the items within the verbal complex.[8]

[Neg > Dat-RWP > Dat-RWP > Acc-RWP > Tense-/Mood-Aux > Adv > Perfective > Adv > Verb]

[Nu	mi	ţi	l-	ar	mai	fi	tot	aruncat]	vrăjitoarea	încoace şi -ncolo.
[not	me$_{dat}$	you$_{dat}$	him$_{acc}$	have$_{cond.3.sg}$	anymore	be$_{perf}$	continually	thrown]	the witch	here and away.

'The witch [wouldn't have kept throwing him] back and forth.'

Figure 1: Preverbal RWPs in [a maximal verbal complex]

Clitic placement RWPs occur in preverbal position with finite (examples 3, 12, 15), infinite (examples 19, 20), and negated imperative verb forms (examples 23, 22), while in postverbal position with participle/gerund (ex. 21) as well as with non-negated imperative verb forms (examples 18, 25, 27).[9] Depending on the formulation, affectionate exclamations, wishes, blessings, and curses can occur in preverbal position (ex. 36) or postverbal position (ex. 37). Due to its phonological shape as /o/, the RWP for 3.sg.acc.f exhibits unique behavior. Preverbally, it occurs only if there is no auxiliary starting with a vowel (ex. 34), otherwise, it occurs postverbally (ex. 35).

[7] In the case of Romanian weak copula verb forms, the prosodic phrase host is the predicative (cf. Section 2.4.2).

[8] The reason why the two dative clitics do not appear in the English translation is because they both can be interpreted as dativus ethicus, which is difficult to express in English. In Romanian, constructions with dativus ethicus are not employed too often nowadays; for their interpretation, see Dobrovie-Sorin and Giurgea (2013:p. 257ff) and Jianu (2013:p. 257ff).

[9] for a comparison to other Romance languages, see Gerlach (2002:p. 267)

Clitic doubling and differential object marking As Spanish, Romanian features clitic doubling, i.e., when an object surfaces twice in specific constructions, both as a clitic and as a full pronoun or noun. In such configurations, the accusative object is marked with the preposition *pe* in Romanian, the counterpart to the Spanish marker *a*. This phenomenon is called *Differential Object Marking* (cf. Tigău 2021).

Combinatorial restrictions As with other Romance languages, RWPs feature arbitrary gaps in clitic-clitic combinations, a phenomenon coined *Person Case Constraint* (cf. Bonet 1994).

Clitic climbing In a few cases, as with the modal *a putea* 'to be able to', Romanian exhibits clitic climbing (ex. 31), where the clitic occurs before the modal instead of the main verb (cf. Monachesi 2005:p. 206).

Coordination As an argument for the difference between proclitics and enclitics, Benincà and Cinque (1993:p. 2323-2324) purport that in Romanian proclitic pronouns can be coordinated, which is not possible with enclitics. However, as mentioned above, RWPs cannot be coordinated, modified or focalized (cf. Dobrovie-Sorin and Giurgea 2013:p. 262), but see the debate on that specific topic presented in Section 1.1. If two coordinated verbs share the same clitic, the clitic has to be repeated for each verb, i.e., the scope of pronominal clitics does not extend across coordination (cf. Monachesi 2001:p. 264).

Interpolation A further syntactic argument for the difference between proclitics and enclitics put forward by Benincà and Cinque (1993:p. 2324-2325) is the interpolation, the occurrence of adverbs or other constituents between a clitic and its host, which is possible in proclitic, but not in enclitic contexts. Yet, to conclude that there is a difference between proclitics and enclitics because of interpolation phenomena is a fallacy: from the fact that the interpolated adverbs have a fixed position, namely right before the main verb, does not follow that there is a difference between proclitic and enclitic pronouns. Given the fix position of the clitic items Dat-RWP>Acc-RWP>AUX both before and after the verbal host and following the same line of reasoning as Benincà and Cinque (1993), one could claim that there is a closer relation between the auxiliary and the main verb in preverbal position (ex. 36) than in postverbal position (ex. 37), where the Acc-RWP occurs between the main and the auxiliary verb. Or that there is a closer relation between an Acc-RWP and the main verb in preverbal position (ex. 26) than in postverbal position (ex. 27), where the Dat-RWP occurs between the main verb and the Acc-RWP. Besides, Pescarini (2019) points to empirical data featuring interpolation phenomena also in enclitic contexts.

2.4. Phonological features of Romanian Weak Pronouns

In order to identify phonological constraints on RWP combinations, one should first categorize the phonological shapes of RWPs based on the different contexts they occur in (for more details, see Avram 1986 or Popescu 2000). Although there is no disagreement about partitioning RWPs into two categories (one for syllabic and one for asyllabic forms), there is a great variety in terminology. Some linguistic RWP descriptions offer a dichotomy between *free* vs. *bound* – as in Iliescu (1975:p. 51), Guţu-Romalo (2008:p. 203), and Dindelegan (2013:p. 382) – between *full* vs. *short* – as in Nastasenco (1997:p. 19) – between *full* vs. *reduced* – as in Popescu (2000:p. 775ff), and Mišeska Tomić (2006:p. 280) – between *full* vs. *non-full* – as in Calude (2001:p. 98) – between *long* vs. *short* – as in Cherecheş (2014:p. 52) – or between *syllabic* vs. *asyllabic* – as in Dobrovie-Sorin and Giurgea (2013:p. 261). Since RWPs can be either syllabic or asyllabic, and these terms adequately and succinctly capture on aspect of the categorical essence of RWPs, I use the terms "*syllabic*" vs. "*asyllabic*".

2.4.1. Î-Prothetic forms

Prothesis, the addition of a sound at the beginning of a word without changing its meaning, is not unusual in Romance languages. In Romanian, *î*-prothesis[10] is encountered with a series of RWPs that have to surface as asyllabic forms and thus need syllabic support, such as *îmi* [ɨm^{i}] or *îți* [ɨts^{i}] (cf. Lombard 1976). In Romanian grammar, these forms are always described as syllabic.

[10]The emergence of Romanian *î*-prothesis occurred in a period between the thirteenth and the sixteenth century (cf. Dindelegan 2016:p. 67).

Notwithstanding the orthography of *î*-prothetic forms, I argue that a more appropriate RWP description abstracts away from the prothetic *î*, classifying these forms as asyllabic. There are three arguments that support this claim.

First, the syllable structure of the *î*-prothetic RWP forms are the same as in other combinations of <syllabic support>-<asyllabic form>, independent of the type of the supporting item. The syllables *îmi* [ɨm^{i}], *și-mi* [ʃim^{i}], *nu-mi* [num^{i}], *că-mi* [cəm^{i}], *să-mi* [səm^{i}], and *dă-mi* [dəm^{i}] have the same structure – <non-pronominal syllabic support> followed by <pronominal asyllabic form> – despite the fact that each syllabic support item is of a different kind: *î* is a prothetic vowel, *și* is the conjunction 'and', *nu* is a negation particle, *că* is a subordinator, *să* is a subjunctive marker, and *dă* is the form |give$_{_2p.sg.imp}$| of the verb *a da* 'to give'. Yet, the asyllabic weak pronoun part, i.e., the string carrying the semantics of the pronoun, is the same in all these combinations, namely *mi* [m^{i}] |cl$_{_1.sg.dat}$|.

Second, the prothetic *î* is required only in very specific contexts: when each of the asyllabic forms *mi* [m^{i}], *ți* [ts^{i}], *i* [i̯], *și* [ʃ^{i}], or *l* [l] occurs alone and there is no vocalic support to the left or to the right. In contexts allowing optionality, the *î*-prothetic forms 'compete' with appropriate syllabic support items (ex. 22 vs. 23 or ex. 42 vs. 43), as described in Section 2.4.3.

Third, *î* is a prothetic vowel not only for weak pronouns but also for weak forms of the verb *a fi* 'to be', such as *îs* [ɨs] in *îs pe munte* 'I am/they are on the mountain' or *îi* [ɨi̯] in *îi acasă* 'he/she is home', as mentioned in Lombard (1976:p. 116), Rosetti (1986:p. 373), or Avram (1986:p. 558). These weak *î*-prothetic verb forms – nowadays used in regional varieties and colloquial language (cf. Avram 1986:p. 558 or Zafiu 2019) – have exactly the same occurrence constraints as the *î*-prothetic RWPs.

These facts demonstrate that treating the prothetic *î* as it is, namely, just prothesis, simplifies the categorization of RWPs by reducing their inventory. Moreover, the abstraction from the prothetic *î* does without Dobrovie-Sorin and Giurgea's (2013:p. 261) **Inside-** vs. **Outside clitic clusters** for dative-syllabic, which is a mixture of form and context. The optimized surface form description of RWPs is presented in Table 1.[11]

2.4.2. Obligatory sandhi

Having clarified the status of *î*-prothetic forms as asyllabic, I now turn my attention to the phonological constraints on RWP combinations to distinguish possible regularities, i.e., patterns, in clitic clusters. As mentioned above, I use the term "sandhi" to refer to phonological cliticization, i.e., the combination of an asyllabic RWP and a syllabic host.

At the beginning of this article, I noted that I use evidence from both historical data and data pertaining to language varieties. As mentioned above in Section 2.4.1, this is actually the case with the Romanian weak verb forms (hereafter RWVs). These forms are *s* [s] – `1.sg/pl.pres` of the verb *a fi* 'to be' with the strong form *sunt*[12] – and *i* [i̯] – `3.sg.pres` with the strong form *e/este*.[13] RWVs can combine with dative RWP forms and can appear both before – *mi-s acolo* in ex. 32 – and after – *acolo mi-s* in ex. 33 – the stressed word *acolo* [a.ˈco.lo] 'there' within the verbal construction (cf. Iliescu 1975:p. 59).

Ignoring any context that enables optional sandhi, which is treated separately in Section 2.4.3, there are two configurations relevant for the description of obligatory sandhi: (1) contexts with no monosyllabic vowel-initial item immediately following the RWP cluster; (2) contexts with an auxiliary starting with a vowel or with the *o*-RWP right after the RWP cluster.[14]

To find out whether there are differences in syllabicity between identical clitic clusters in pre- vs. post-verbal contexts, each RWP combination has to be taken into account. In the case of RWP-RWV combinations, it is about pre- vs. post-predicative contexts. Due to the parameters `RIGID ORDER` and `MAXIMAL NUMBER`, the following possible configurations have to be examined: a single Acc-RWP, a single Dat-RWP, the sequence Dat-RWP>Acc-RWP, the sequence Dat-RWP>Dat-RWP, and the sequence Dat-RWP>Dat-RWP>Acc-RWP. Moreover, due to the similar behavior between RWVs and *î*-prothetic RWPs, contexts with a single RWV, the sequence Dat-RWP>RWV, and the sequence Dat-RWP>Dat-RWP>RWV are relevant, too.

[11] for similar synopses, see (Avram 1986:p. 554) or Dobrovie-Sorin and Giurgea (2013:p. 261)

[12] spelled *sînt* before the orthographic rules adopted in 1993 by the Romanian Academy

[13] There are other weak verb forms – such as the future auxiliary for `2p.sg` *îi* and for `2p.pl` *îți* (cf. Avram 1986:p. 558) – which can be modeled on a par with the present weak verb forms treated here. These future auxiliary forms are homonymous with the RWPs *îi* and *îți*, respectively.

[14] Since the *o*-RWP behaves like a vowel-initial auxiliary and is treated accordingly, references to an RWP cluster in this study do not include this item.

Case	*Number*	*Person*	*Type*	*Gender*	*Syllabic*	*Asyllabic*	
						onset	***coda***
A	*Sg*	*1.*	*pers/refl*	*m/f*	mă [mə]	m [m]	—
c		*2.*	*pers/refl*	*m/f*	te [te]	te [te̯]	—
c		*3.*	*pers*	*m*	—	l [l]	l [l]
u				*f*	o [o]	o [o̯]	—
s			*relf*	*m/f*	se [se]	se [se̯], s [s]	—
a	*Pl*	*1.*	*pers/refl*	*m/f*	ne [ne]	ne [ne̯]	—
t		*2.*	*pers/refl*	*m/f*	vă [və]	v [v]	—
i		*3.*	*pers*	*m*	—	i [i̯]	i [i̯]
v				*f*	le [le]	le [le̯]	—
e			*relf*	*m/f*	se [se]	se [se̯], s [s]	—
	Sg	*1.*	*pers/refl*	*m/f*	mi [mi]	mi [mi̯]	mi [m^{i}]
D		*2.*	*pers/refl*	*m/f*	ți [tsi]	ți [tsi̯]	ți [ts^{i}]
a		*3.*	*pers*	*m/f*	i [i]	i [i̯]	i [i̯]
t			*relf*	*m/f*	și [ʃi]	și [ʃi̯]	și [ʃ^{i}]
i	*Pl*	*1.*	*pers/refl*	*m/f*	ni [ni], ne [ne]	ni [ni̯], ne [ne̯]	—
v		*2.*	*pers/refl*	*m/f*	vi [vi], vă [və]	vi [vi̯], v [v]	—
e		*3.*	*pers*	*m/f*	li [li], le [le]	li [li̯], le [le̯]	—
			relf	*m/f*	și [ʃi]	și [ʃi̯]	și [ʃ^{i}]

Table 1: **Surface forms of Romanian weak pronouns**: orthographic form [IPA]

I. Contexts without a monosyllabic vowel-initial item following the RWP cluster

A. RWPs/RWVs in the right-most position

A1. consonants: Always asyllabic, *l*-RWP and *s*-RWV occur only in the right-most position.

A2. *i*-forms: All dative singular *i*-RWPs, the accusative plural *i*-RWP as well as the *i*-RWV occur in the right-most position as asyllabic forms, either as palatalized consonants[15] (ex. 40) or as glides (ex. 42).

A3. *e*- and *ă*-forms: All RWP forms ending in *ă* [ə] or *e* [e] occur in the right-most position as syllabic forms (ex. 12). The plural dative RWPs surface as dative-accusative case-syncretic forms,[16] as in *vă dă mere* [və.ˈdə.ˈme.re] |cl$_{\text{2.pl.dat}}$ give$_{\text{3.sg.pres}}$ apple$_{\text{acc.pl.indf}}$| 'he/she gives you apples'.

[15] In her prosodic analysis of RWPs, Chereches¸ (2014) observes a parallel between RWP clusters with *i*-forms and, e.g., plural masculine nouns ending in *i*: "the plural marker for masculine nouns -i reduces to a palatalization gesture word-finally but stays a full vowel when followed by extra inflectional material" (op. cit. p. 55), for instance, *lup* [ˈlup] 'wolf' vs. *lupi* [ˈlup^{i}] 'wolves' vs. *lupilor* [ˈlu.pi.lor] 'of/to the wolves'. Avram (1986:p. 552) refers to this palatalization gesture as 'final pseudo-i'. In the context of RWPs, the right-most position in a cluster is equal to word finality in Chereches¸' (2014) example.

[16] According to Giurgea (2013), the RWP forms for plural dative such as *ni*, *vi*, and *li* emerged by a kind of dissimilation process, in analogy to the singular dative *i*-forms such as *mi* or *ți*.

B. RWPs in non-right-most positions

In all other positions, only syllabic *i*-vocalic dative RWP forms can occur (ex. 26; cf. footnote 15).

C. Clitic-host relation

C1. single asyllabic forms: For RWP/RWV items that surface as single asyllabic forms, the use of *î*-prothesis as syllabic support is obligatory in pre-verbal/pre-predicative position. In post-verbal position, the verb is the syllabic support for the asyllabic RWPs. In post-predicative position, both the prothetic *î* and the predicate can be syllabic host to an RWV.

C2. syllabic forms: For RWPs that surface as syllabic forms – alone or in a cluster – there is no need for syllabic support: both pre- and post-verbally, they occur as independent, unstressed, syllabic forms.

C3. clusters: In other contexts, i.e., in RWP/RWP-RWV clusters with two or three items, the asyllabic right-most item uses its neighbor to the left as syllabic support.

Overall picture: Any right-most obligatory asyllabic item – alone or in an RWP/RWP-RWV cluster – has its *syllabic host to the left*. The host can be an *î*-prothetic vowel (ex. 3), a syllabic RWP (ex. 24 or ex. 33), or a verb (ex. 18). Hence, in this context, all asyllabic items are enclitics at the syllabic level, independent of whether they are proclitics or enclitics at the phrasal level.

II. Contexts with monosyllabic vowel-initial items following the RWP cluster

In contexts with monosyllabic vowel-initial items in the verbal complex – such as the *o*-RWP or the auxiliary *ar* – RWV items do not occur (an accusative object is incompatible with a construction of the verb *a fi* 'to be'). As already mentioned in the syntax sketch, the *o*-RWP is an exception. If both the *o*-RWP and a vowel-initial monosyllabic auxiliary have to be expressed in the verbal complex, the *o*-RWP always occurs postverbally as a syllable (ex. 34 vs. 35; see the phonological constraint `NO HIATUS` in Popescu 2000 or Chereches¸ 2014). In seldom contexts, when both the *o*-RWP and a vowel-initial monosyllabic auxiliary occur postverbally, the *o*-RWP forms a diphthong to the following vowel as a glide, thus becoming an asyllabic proclitic just like any other RWP, as in ex. 38 (cf. also Bošković 2001:p. 123, footnote 27).

A. RWPs in the right-most position

A1. consonants: In addition to the *l*-RWP, the *ă*-forms *mă* and *vă* as well as the *se*-RWP surface in this context as [m], [v], and [s], respectively (*m-ai văzut* [ma^{i}.və.^{l}zut] |cl$_{_1.sg.acc}$ have$_{_2.sg.pres}$ see$_{_part.perf}$| 'you have seen me').

A2. *e*-forms: With the exception of the reflexive *se* (see above), all accusative and dative forms ending in *e* are obligatorily asyllabic, featuring the glide [e̯] (ex. 36).

A3. *i*-forms: All *i*-RWP forms are obligatorily asyllabic, featuring the glide [i̯] (ex. 34).

B. RWPs in non-right-most positions

In all other positions, there is no change to the previous context type, and only syllabic *i*-RWP forms in dative can occur (cf. footnote 15).

C. Clitic-host relation

All items in the right-most position surface without exception as asyllabic items, thus needing syllabic support, which, in this context, is provided by the following item. The glides form a diphthong with the following vowel, while the consonants are onsets of the syllabic host.

Overall picture: Any right-most item – alone or in an RWP cluster – is an obligatorily asyllabic item and has its *syllabic host to the right*. The host can be the *o*-RWP (ex. 34) or any monosyllabic vowel-initial item (examples 35, 36, 37, 39). Hence, in this context, all asyllabic items are proclitics at the syllabic level, independent of their position towards the verb.

This RWP analysis of obligatory sandhi indicates that (1) the RWP system is not idiosyncratic, and (2) there is no difference in syllabicity between RWP clusters in pre- vs. post-verbal position. Thus, the claim that clitics occupying postverbal position "show obligatory phonological cliticization" (Dobrovie-Sorin 1999a:p. 533) is untenable. The same applies to the claim that "[p]ostverbal weak pronouns always encliticise" (Klein 2007:p. 62), where "the process of cliticisation is a phonological process" (Klein 2007:p. 61). Neither Dobrovie-Sorin (1999a) nor Klein (2007) gives an accurate description of the differences in phonology between RWPs in pre- vs. post-verbal positions, e.g., by showing which is the *phonological host* and which the *phonological clitic* in the two environments. In particular, in ex. 26 and ex. 27, there is no phonological cliticization involved whatsoever, neither preverbally in ex. 26 nor postverbally in ex. 27: on both sides of the verb, all RWPs are syllabic. This is also illustrated by the correct analysis of RWP syllabicity for exampes 12, 13, 15 and 16 as well as the analysis of the hyphen used only as a postverbality marker in the "Disambiguation" column for ex. 13 and 16 in Appendix 1. It is obvious that both Dobrovie-Sorin (1999a) and Klein (2007) overlook a very subtle ambiguity in Romanian orthography, namely, the *hyphen ambiguity* (see Section 2.5).

As a matter of fact, it is not the clitic sequences that differ in pre- vs. post-verbal position, but some verb forms that have to adjust for specific enclitic configurations. The Romanian gerund without enclitics ends in *-nd* as in *dând mere* [ˈdɨnd.ˈme.re] |give$_{_ger}$ apple$_{_acc.pl.indf}$| 'giving apples'. This ending doesn't change when followed by the *o*-RWP: *dând-o* [ˈdɨn.do] |give$_{_ger}$ cl$_{_3.sg.acc.f}$| 'giving it/her'. However, when followed by other clitics or clitic sequences, the gerund form features a final *-u* as in ex. 21 (cf. Maiden et al. 2021:p. 149). The same is the case with verb forms ending in an asyllabic *-i* (cf. Footnote 15), which becomes syllabic when followed by enclitics: *Le dați afară.* [le.ˈdatsʲ.a.ˈfa.rə] |cl$_{_3.pl.acc.f}$ throw$_{_2.pl.pres}$ out| 'You throw them out.' vs. *Dați-le afară!* [ˈda.tsi.le.a.ˈfa.rə] |throw$_{_2.pl.imp}$ cl$_{_3.pl.acc.f}$ out| 'Throw them out!'. And again, the *o*-RWP is an exception thereof: *Dați-o afară!* [ˈda.tsi̯o.a.ˈfa.rə] |throw$_{_2.pl.imp}$ cl$_{_3.sg.acc.f}$ out| 'Throw it/her out!'. Accordingly, *u*-epenthesis phenomena in the given contexts are instances of obligatory sandhi of gerund forms.

2.4.3. Optional sandhi

When the context is favorable, optional sandhi in RWPs may emerge in contexts complementary to the obligatory sandhi ones.

A. RWPs in preverbal position

A1.: when the left-most item before a single asyllabic RWP ends with a vowel (ex. 40 vs. 41);

A2.: when the verb begins with an unstressed vowel right after a single asyllabic RWP (ex. 42 vs. 43);

A3.: when the verb begins with an unstressed vowel immediately after the right-most syllabic RWP – *le aduci* [le.a.ˈdutʃʲ] vs. *le-aduci* [le̯a.ˈdutʃʲ] |cl$_{_3.pl.acc.f}$ bring$_{_2.sg.pres}$| 'you bring them' or *mi le aduci* [mi.le.a.ˈdutʃʲ] vs. *mi le-aduci* [mi.le̯a.ˈdutʃʲ] |cl$_{_1.sg.dat}$ cl$_{_3.pl.acc.f}$ bring$_{_2.sg.pres}$| 'you bring them to me'.

Both **A1.** and **A2.** and the combination thereof – such as *că îmi aduci mere* [cə.ɨmʲ.a.ˈdutʃʲ.ˈme.re] |that î$_{_host}$cl$_{_1.sg.dat}$ bring$_{_2.sg.pres}$ apple$_{_acc.pl.indf}$| 'that you bring me apples' – illustrate the 'competition' between *î*-prothetic vowels and other contextually appropriate syllabic support items, as mentioned in Section 2.4.1. Strictly speaking, these are instances of obligatory asyllabic clitics with optional choice of syllabic host.[17] By contrast, **A3.** illustrates genuine optional sandhi, i.e, contexts where a syllabic RWP form 'competes' with its asyllabic counterpart: [le] vs. [le̯].

In the construction Dat-RWP>TO-BE, the plural dative *i*-forms combine with the verb as *i*-glides in optional sandhi, i.e., on a par with the singular dative *i*-RWP forms, as in *li-e sete* [li̯e.ˈse.te] vs. *le e sete* [le.i̯e.ˈse.te] |cl$_{_3.pl.dat}$ be$_{_3.sg.pres}$ thirst| 'they are thirsty'.

[17] Mutatis mutandis, this is the case with RWV forms in similar contexts, too.

B. RWPs in postverbal position

The only configuration that allows RWP optional sandhi in postverbal position is when a word begins with an unstressed vowel directly after the the right-most syllabic RWP (ex. 27 vs. 28).

C. Clitic-host relation

C1. [asyllabic RWP = clitic]

A single asyllabic RWP can have a host both to its left, e.g., a subordinator (ex. 40) or the prothetic *î* (ex. 41) and to its right, e.g., the main verb (ex. 42).

A right-most syllabic RWP can combine in optional sandhi only to its right: preverbally with the verb – *le-aduci* [lḛa.ˈdutʃʲ] |cl$_{\text{3.pl.acc.f}}$ bring$_{\text{2.sg.pres}}$| 'You bring them.' – or postverbally with an item with non-stressed initial vowel that immediately follows the VERB-RWPs sequence – *adu-le-acolo* [a.ˈdu.lḛa.ˈco.lo] |bring$_{\text{2.sg.imp}}$ cl$_{\text{3.pl.acc.f}}$ there| 'Bring them there!'.

C2. [syllabic RWP = host]

Due to its sonority, the vowel *î* [ɨ] at the beginning of a word – other than RWPs with *î*-prothesis – is deleted in optional sandhi. This allows both for optional sandhi on each side of a monosyllabic vowel-only item such as **o**-RWP – *s-o-ncep* [son.ˈtʃep] vs. *s-o încep* [so.ɨn.ˈtʃep] vs. *să o încep* [sə.o.ɨn.ˈtʃep] |that cl$_{\text{3.sg.acc.f}}$ start$_{\text{1.sg.pres}}$| 'that I start it' – and for a co-occurrence of obligatory and optional sandhi in the same context – *le-o-ntind* [lḛon.ˈtind] vs. *le-o întind* [lḛo.ɨn.ˈtind] |cl$_{\text{3.pl.dat}}$ cl$_{\text{3.sg.acc.f}}$ stretch$_{\text{1.sg.pres}}$| 'I stretch it for them'. Interestingly, in such contexts, even a single, otherwise obligatory asyllabic RWP such as *mi* [mʲ] can surface as syllabic form *mi* [mi], replacing the initial central vowel *î* [ɨ] of the verb – ex. 44[18] vs. 45.

Such an instance of **interlocked cliticization** between the phrasal and the syllabic level is detailed in ex. 29 (vs. ex. 30) – *se-ntâmplă* 'it happens'. Here, the monosyllabic reflexive RWP *se* is the syllabic host for the asyllabic segment [n] of the verb *întâmplă*, while, at the same time, the verb itself is the phrasal host for the monosyllabic, non-stressed RWP *se*.

Stressed vowel-initial items do not allow for optional sandhi, neither preverbally – **le-aflu* vs. *le aflu* [le.ˈa.flu] |cl$_{\text{3.pl.acc}}$ find$_{\text{1.sg.pres}}$| 'I find them' – nor postverbally – **Dă-le-altuia!* vs. *Dă-le altuia!* [ˈdə.le.ˈal.tu.i̯a] |give$_{\text{2.sg.imp}}$ cl$_{\text{3.pl.acc.f}}$ other$_{\text{sg.dat.m}}$| 'Give them to another!'. This seems to be related to the fact that RWPs cannot be stressed. However, there are some contexts of optional sandhi where the stressed negation particle *nu* loses the vowel, while the syllabic host – the immediately following monosyllabic vowel-initial item – acquires the stress, as in *n-o văd* [ˈno.ˈvəd] vs. *nu o văd* [ˈnu.o.ˈvəd] |NEG cl$_{\text{3.sg.acc.f}}$ see$_{\text{1.sg.pres}}$| 'I don't see her/it' or *n-am văzut-o* [ˈnam.və.ˈzu.to] vs. *nu am văzut-o* [ˈnu.am.və.ˈzu.to] |NEG have$_{\text{1.sg.pres}}$ see$_{\text{part.perf}}$ cl$_{\text{3.sg.acc.f}}$| 'I haven't seen her/it'.

Note that optional sandhi between shorter – usually monosyllabic – items is much more prevalent than between a monosyllabic and a heavy polysyllabic item (cf. also Gerlach 2002:p. 141). That means that, if there is a choice for optional sandhi between two items with different weights, the combination to the shorter one might be preferred.[19]

[18] Original as *roua dimineții mi-mbată inima* 'the morning dew makes my heart drunk' retrieved from the URL https://poeziipentrusufletulmeu.com/2019/09/07/dorule/ on 2022-02-27, but, since *dimineții* is not an essential part of the example, I have left it out for reasons of space.

[19] Although further discussion of this topic would go beyond the scope of the current study, an intriguing question concerns the contexts in which optional sandhi is chosen. When does a speaker decide on a variant with optional sandhi and when not? Is optional sandhi truly optional, or are there some conditions involved that we have yet to recognize? There are contradicting opinions on this topic, e.g., Popescu (2003), claiming that the trigger for optional sandhi is **speech rate** vs. Dindelegan (2013:p. 388), claiming that optional sandhi is controlled by **language register rules**.

2.5. Orthography issues

Writing is the most prominent medium of communication in science. Hence, writing and notation systems, both for natural language and for specific scientific domains, are crucial for a correct understanding of messages, ideas, and argumentation. Yet these systems are not perfect, and they are prone to changes and improvements. The orthography of a language can be designed with a phonetic or an etymological principle in mind, and can be easier or more difficult to master, even by native speakers (cf. Firică 2009). The Romanian orthographic system is no exception.

2.5.1. Ambiguities

Romanian orthography exhibits various types of ambiguity, which hinder an easy understanding of the RWP data, such as different types of homonymy as well as hyphen ambiguity.

For instance, there is homonymy[20] – actually, both homophony and homography – between the indefinite article `fem.sg.nom-acc` *o*, the cardinal numeral feminine *o*, the future particle *o*,[21] and the RWP *o*, as in *o comisie o să o vadă numai o zi* [o.co.ˈmi.si.e.o.sə.o.ˈva.də.ˈnu.maɨ̯.o.ˈzi] |o$_{\text{art.sg.indf.f}}$ committee o$_{\text{part.fut}}$ that cl$_{\text{3.sg.acc.f}}$ see$_{\text{3.sg.conj.pres}}$ only o$_{\text{card.num.f}}$ day| ‘a committee will see her only one day’. Another instance of the same type is the dative-reflexive RWP *și* and the conjunction *și* ‘and’, as in *și le cumpără* ***și*** *și le revinde* [ʃi.le.ˈcum.pə.rə.ʃi.ʃi.le.re.ˈvin.de] |cl$_{\text{3.sg.dat.refl}}$ cl$_{\text{3.pl.acc.f}}$ buy$_{\text{3.sg.pres}}$ and$_{\text{conj}}$ cl$_{\text{3.sg.dat.refl}}$ cl$_{\text{3.pl.acc.f}}$ resell$_{\text{3.sg.pres}}$| ‘he/she buys them for him-/herself **and** resells them for him-/herself’.

Grapheme-phoneme ambiguity is evidenced, e.g., by the orthographic form *mi*-RWP for `1p.sg.dat`, which can be either the syllabic form [mi] (ex. 24), the asyllabic form with a glide [mɨ̯] (ex. 34), or the asyllabic palatalized form [mʲ] (ex. 40).

There is a particularly treacherous homonymy between the *ți*-RWP and the imperative plural suffix *ți* as in *Pune-ți-l jos!* [ˈpu.ne.tsil.ˈ os] |put$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.dat}}$ cl$_{\text{3.sg.acc.m}}$ down| ‘Put yours down!’ (said to a single person) vs. *Puneți-l jos!* [ˈpu.ne.tsil.ˈ os] |put$_{\text{2.pl.imp}}$ cl$_{\text{3.sg.acc.m}}$ down| ‘Put it down!’ (said to a group of people).

A very subtle ambiguity in the current standard Romanian orthography concerns the use of hyphen: among other, the hyphen is used as sandhi marker, postverbal marker, or both (cf. also Dobrovie-Sorin and Giurgea 2013:p. 262). Appendix 1 shows some instances of hyphen usage and how it corresponds to different clitic-host structures and relations.

2.5.2. Postverbal marking

Interestingly, many languages featuring clitics employ special orthographic marking for enclitics, but not for proclitics. What is the reason for this unequal treatment? A possible explanation is the reduction of cognitive load in language processing, more precisely, the **reduction of extraneous (or extrinsic) cognitive load**, the cognitive load resulting from the way in which something is presented (cf. Sweller et al. 2019).

Due to the different positions relative to the verb and the linearity of the utterance, clitics between two verbs can be interpreted either as enclitic to the preceding verb or as proclitic to the subsequent verb. Different attempts at reducing potential ambiguities are possible: in speech, this is done with the contour of the prosodic phrase, while in writing, different specific orthographic rules are used. The postverbality of an enclitic sequence is explicitly marked in Romanian by a hyphen between the preceding verb and the clitic sequence, while in Italian and Spanish the verb and the clitic sequence are written together as a single orthographic unit.[22]

Example 1[23] and 2 illustrate my assessment of this issue. In ex. 1a, the hyphen links the noun *prietena* ‘girl-friend’ to its post-nominal possessive clitic *mi* ‘my’, an rarely used possessive construction in modern Romanian. The possessive enclitic must be evaluated with respect to the preceding noun phrase, not to the

[20] see also footnote 13

[21] used in colloquial language (cf. Dragomirescu et al. 2022:p. 245-246)

[22] This is essentially the same as using brackets in mathematics to clearly mark the scope of individual operators.

[23] adapted from Avram (1986:p. 561)

following verbal phrase.[24] In ex. 1b, the dative clitic *mi* and the accusative clitic *le* are both proclitics to the verb, and thus no marking is needed or even allowed. Ex. 2 features a similar problem in Spanish: the parsing and understanding of a series of enclitic and proclitic pronouns is eased by writing the first verb *hablar* and its enclitic *le* together as one orthographic unit – *hablarle* – so that it should not be evaluated as proclitic to the following verb, as it is the case with *lo*.

(1) a. prietena-mi le dă mere
girlfriend-my to_them give apples
'My girlfriend gives them apples.'

b. prietena mi le dă înapoi
girl-friend to_me them give back
'The girlfriend gives them back to me.'

(2) Al hablarle lo detesté.
To_the talking to_him him hated.
'When I talked to him, I hated him.'

Based on psycho-linguistic experiments on Italian single enclitics described in Finocchiaro and Caramazza (2006), Finocchiaro (2005) ponders whether the interpretation of the experiment results, namely, that enclitics pattern as affixes do with respect to gender-congruence, can be safely extended to proclitics, which are written separately from the verb, so they are "orthographically independent" (e.g., Ital. *lo pórto* |lo$_{\text{_cl.3.sg.m}}$ portare$_{\text{_1.sg.pres}}$| 'I bring it'). For Finocchiaro (2005),

> "[…] asymmetries between enclitics and proclitics are well known and appear to extend beyond superficial graphical differences. Specifically, the relation between the proclitic and the host verb appears to be less strong than the relation between the enclitic and the host verb (Benincà and Cinque 1993). Benincà and Cinque (1993) argued that the graphical difference between enclitics and proclitics corresponds to deep structural differences." (op. cit. p. 303)

However, Luraghi (2017) mentions Benincà & Cinque's (1983)[25] suggestion that

> "the fact that enclitics are often attached to their host graphically whereas proclitics are not may reflect some difference in the relation between the host and the clitic based on the direction of liaison" (op. cit. p. 189),

which is a rather reasonable suggestion.

Undeniably, there *are* differences between proclitics and enclitics due to the linearity of the utterance, the relative position to the verb as well as to the fact that both proclitics and enclitics have to be interpreted as parts of the same verbal phrase. Yet, whether differences in the orthographic rules of a language at an arbitrary point in time can be linked to deep structural differences motivated by some theory-internal assumptions is questionable.

It is useful to evaluate this claim in a more general, comparative context: in Italian, the verb and the enclitic are written together; in Romanian, they are linked by a hyphen; in Bulgarian, however, there is no difference in the orthographic representation of proclitics vs. enclitics – both are separated from the verb by a space. These observations could be interpreted to imply that in Italian, there is a stronger link between verb and enclitic than in Romanian, where there is a weaker link, and, in turn, that in Bulgarian there is the weakest or even no link between verb and enclitic. However, this is not a valid interpretation because orthographic rules are ultimately language-specific and do not reliably represent morphophonological relationships in a consistent way across languages. Indeed, every now and then they are even subject to change, independent of actual linguistic change.[26]

[24] For RWP-RWV clusters after a prosodic host, post-predicative marking is not used, as illustrated by ex. 33, where there is no hyphen between *acolo* and *mi-s*.

[25] Obviously, the mention "Benincà & Cinque (1983)" in Luraghi (2017) is a typo of the year of Benincà and Cinque (1993).

[26] see, e.g., Johnson (2005) for changes in the spelling of compounds introduced by a reform of German orthography in 1996

3. From data to model

In the previous section, I presented a detailed analysis of RWPs and RWVs as they occur in specific contexts, i.e., a static view of the data. A computational linguistic model for generating correct RWP surface forms in appropriate contexts, i.e., a procedural view, implies a set of input items (i.e., the underlying forms) and a set of rules that transform each input into the corresponding output. To construct such a model, I try to grasp patterns in language and express them in a formal way. Yet, since language is in steady change, some phenomena can be difficult to make out only by considering a synchronic perspective: the data might be somehow incomplete and/or idiosyncratic. It is essentially like looking at a painting from too close up: you can see tiny details, but not the whole picture.

By taking a step back in time as well as a step aside to some closely related language such as Aromanian, it is possible to find missing pieces to the RWP puzzle. Indeed, in the literature on Old Romanian, there is evidence for the existence of forms such as *su* and *lu* (cf. Graur 1960, Avram 1986:p. 652, or Dindelegan 2016). Consider, e.g., the instances *eu measeru-su* 'I am poor' in Dindelegan (2016:p. 169) or *nu vrea de să-lu ştie* 'he does not want anybody to know him' in Dindelegan (2016:p. 242). Moreover, a comparison between Aromanian and Romanian shows that the Aromanian form *lu* is the counterpart of the – now consonant-only – Romanian form *l*, as illustrated in Figure 2 (reproduced from Marioţeanu 1994:p. 14-15).

Figure 2: Aromanian *lu* vs. Romanian *l* in Marioţeanu (1994:p. 14-15)

As Krämer (2012) aptly notes, '[t]he existence, status and form of underlying representations have been hotly debated in phonological research', hence, it is difficult to agree on this kind of abstractions. However, the underlying representations proposed here provide a much better justification for the linguistic reality than, for instance, Dobrovie-Sorin and Giurgea's (2013:p. 266) assumption that "clitic forms are underlyingly asyllabic or syllabic", or Popescu's (2003:p. 154) assumption of unspecified underlying mora. Why is this the case? The model proposed here employs input entities evidenced in the history of Romanian and also in the closely related Aromanian, namely, items such as *su* and *lu*. Moreover, since it treats both Romanian weak pronouns and Romanian weak verbs uniformly, it offers a broader coverage of the modeled phenomena.

With this in mind, the set of *underlying items* for the model proposed here can now be established: all input items are syllabic (cf. Table 2). The constraints ruling the generation of appropriate surface forms can be derived from the description of RWP sandhi. For consonant-only RWPs/RWVs such as *l* or *s*, the model takes corresponding syllabic input[27] – *lu* or *su* – and applies the constraints derived from the data analysis. The right-most item containing the vowel *u* or *i* becomes asyllabic: *u* is deleted – cf. the optional sandhi for negation *nu* (*nu o văd* vs. *n-o văd* 'I don't see her') – while *i* becomes a palatalization gesture or a glide, depending on the context. Since the accusative always occurs after the dative, hence always in the right-most position, the syllabic input *lu* always surfaces as asyllabic, hence as a consonant. The same applies to the syllabic input *su* for the *s*-RWV forms.

This model can be implemented in computational linguistics as well. For instance, in Gerstenberger (2018), I sketch the constraints required for a computational-linguistic generation of correct RWP forms for a given context which, in turn, is couched into a general framework for linearization, the *General Linearization Model*, as proposed in Gerstenberger (2007). In this, the goal is to compare the results of this relatively simple rule-based model to the output of different statistical-based models.

[27] Already more than 60 years ago, Graur (1960:p. 847) hinted at the possibility of modeling the asyllabic *l*-RWP that way, namely using the syllabic *lu* as underlying representation.

Number	**Accusative**					**Dative**					**Verb** *a fi*	
	1p	2p	3p.m	3p.f	3p.refl	1p	2p	3p.m	3p.f	3p.refl	1p.pres	3p.pres
Sg	/mə/	/te/	/lu/	/o/	/se/	/mi/	/tsi/	/i/	/i/	/ʃi/	/su/	/i/
Pl	/ne/	/və/	/i/	/le/	/se/	/ni/	/vi/	/li/	/li/	/ʃi/	–	/su/

Table 2: Input for the surface form generation of Romanian weak pronouns and weak verb forms

4. Conclusion

In this study, I have presented an analysis of Romanian weak pronouns based on two orthogonal levels: a plain phrasal level – with the stressed verb as phrasal host and unstressed weak pronouns as phrasal clitics – and an intricate syllabic level – with the syllabic item as 'host' and the asyllabic item as 'clitic'.

Unlike the traditional descriptions hitherto, which unanimously classify pronominal *î*-prothetic forms as syllabic, I have used empirical evidence to argue for abstracting away from the prothetic *î*, and instead, classifying such forms as asyllabic. I sketched the syntactic configurations of the RWPs as well as the surface forms in which these items must, in the case of *obligatory sandhi*, or may, in the case of *optional sandhi*, occur. Through a careful examination of the orthography employed to represent RWPs, I identified different types of ambiguities that have led to an inaccurate description of syllabic postverbal RWP instances as phonological clitics by both Dobrovie-Sorin (1999a:p. 533) and Klein (2007:p. 62).

Since language is perpetually in a state of flux, it is not always possible to build a regular model for specific phenomena only from a synchronic perspective; this is made more difficult by the fuzziness of the concept 'synchronic' in terms of time frame delimitation in language description. Given this circumstance, expanding the view of the language data both historically and concerning language varieties and closely related languages such as Aromanian, I found evidence that leads to a model for RWP surface form generation without the idiosyncrasies asserted by Barbu (1999), without the assumption of unspecified underlying mora for *î*-protetic forms as in Popescu (2003:p. 154), without Klein's (2007:p. 77) employment of clusters of *î*-prothetic forms as model input, without Chereches¸' (2014:p. 56) issues with asyllabic consonantal forms lacking underlying vowels, and without Dobrovie-Sorin and Giurgea's (2013:p. 266) assumption of mixed asyllabic and syllabic underlying representations.

Finally, by providing substantiated evidence from elaborate data analyses, I have argued against a dissimilar treatment of weak pronouns occurring in preverbal as opposed to postverbal position. Hence, my answer to the question posed in the title "*How weak are Romanian clitic pronouns?*" is as follows. Since there is no crucial difference between proclitics and enclitics, there is no reason to make a distinction between weak pronouns and clitic pronouns in Romanian either.

Acknowledgements

I wish to thank Trond Trosterud and Laura Janda for their constructive discussions on this topic as well as Joshua Wilbur and Øystein Vangsnes for their thoughtful and detailed feedback on this article. Moreover, I would like to thank two anonymous reviewers for their valuable comments and suggestions on an earlier draft. Of course the usual disclaimers apply.

References

Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199279906.001.0001.

Avram, Andrei. 1986. Sandhi Phenomena in Romanian. In *Sandhi Phenomena in the Languages of Europe*, edited by Henning Andersen, pp. 551–574. Mouton de Gruyter. https://doi.org/10.1515/9783110858532.551.

Avram, Mioara. 1997. *Gramatica pentru toţi*. Editura Humanitas, Bucureşti.

Barbu, Ana Maria. 1999. Complexul verbal. *Studii şi Cercetări Lingvistice* pp. 39–84. Available at http://dspace.bcu-iasi.ro/handle/123456789/8747.

Barbu, Roxana-Maria and Ida Toivonen. 2018. Romanian Object Clitics: Grammaticalization, agreement and lexical splits. In *Proceedings of the LFG18 Conference*, edited by Miriam Butt and Tracy Holloway King, p. 67–87. Stanford: CSLI Publications.

Benincà, Paola and Guglielmo Cinque. 1993. Su alcune differenze tra enclisi e proclisi. In *Omaggio a Gianfranco Folena (vol. 3)*, pp. 2313–2326. Editoriale Programme, Padova.

Blaszczak, Joanna, Dorota Klimek-Jankowska, and Krzysztof Migdalski (eds.) . 2015. *How Categorical are Categories?: New Approaches to the Old Questions of Noun, Verb, and Adjective*. De Gruyter Mouton. https://doi.org/10.1515/9781614514510.

Bonet, Eulàlia. 1994. The person-case constraint: A morphological approach. *MIT Working Papers in Linguistics* 0 22: 33–52.

Bonet, Eulàlia M. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Ph.D. thesis, MIT. Available at http://hdl.handle.net/1721.1/13534.

Bošković, Željko. 2001. *On the Nature of the Syntax-Phonology Interface*. Emerald Group Publishing. https://doi.org/10.1163/9780585474250.

Bredemeier, Jürgen. 1976. *Strukturbeschränkungen im Rumänischen. Studien zur Syntax der prä- und postverbalen Pronomina*. TBL Verlag Gunter Narr, Tübingen.

Brown, Keith (ed.) . 2006. *Encyclopedia of Language and Linguistics*. Elsevier. Available at https://www.elsevier.com/books/encyclopedia-of-language-and-linguistics/brown/978-0-08-044299-0.

Caink, Andrew D. 2006. Clitics. In Brown (2006), pp. 491–495. https://doi.org/10.1016/B0-08-044854-2/00110-3. Available at https://www.elsevier.com/books/encyclopedia-of-language-and-linguistics/brown/978-0-08-044299-0.

Calude, Andreea S. 2001. Romanian clitics: Siding with the Serbo-Croatian or the French? *Revue roumaine de linguistique* 46 1-4: 91–104. Available at https://www.calude.net/andreea/Research.html.

Cardinaletti, Anna. 1999. Pronouns in Germanic and Romance Languages: An overview. In van Riemsdijk (1999b). https://doi.org/10.1515/9783110804010.

Cardinaletti, Anna and Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In van Riemsdijk (1999b). https://doi.org/10.1515/9783110804010.

Chereches, Anca. 2014. A Prosodic Analysis of Romanian Pronominal Clitics. In *University of Pennsylvania Working Papers in Linguistics*, vol. 20. Available at https://repository.upenn.edu/pwpl/vol20/iss1/7/.

Di Sciullo, Anna Maria and Edwin Williams. 1987. *On the Definition of Word*. MIT Press, Cambridge (Mass.). Available at https://mitpress.mit.edu/books/definition-word.

Dindelegan, Gabriela Pană (ed.) . 2013. *The Grammar of Romanian*. Oxford University Press. Available at https://global.oup.com/academic/product/the-grammar-of-romanian-9780199644926.

Dindelegan, Gabriela Pană (ed.) . 2016. *The Syntax of Old Romanian*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198712350.001.0001.

Dindelegan, Gabriela Pană, Adnana Boioc Apintei, and Blanca Croitor (eds.) . 2019. *Variație diacronică și diatopică: Note gramaticale*. Editura Universităţii din Bucureşti. Available at https://editura-unibuc.ro/en/magazin/filologie/limba-romana/variatie-diacronica-si-diatopica-note-gramaticale/.

Dobrovie-Sorin, Carmen. 1994. *The Syntax of Romanian: Comparative Studies in Romance*. Mouton de Gruyter. https://doi.org/10.1515/9783110886597.

Dobrovie-Sorin, Carmen. 1999a. Clitics across categories: The case of Romanian. In van Riemsdijk (1999b).

https://doi.org/10.1515/9783110804010.
Dobrovie-Sorin, Carmen. 1999b. The typology of pronouns and the distinction between syntax and morphophonology. In van Riemsdijk (1999b). https://doi.org/10.1515/9783110804010.
Dobrovie-Sorin, Carmen and Ion Giurgea (eds.) . 2013. *A Reference Grammar of Romanian*, vol. 1: The noun phrase. John Benjamins. https://doi.org/10.1075/la.207.
Dragomirescu, Adina, Alexandru Nicolae, and Rodica Zafiu. 2022. The loss of analyticity in the history of Romanian verbal morphology. In Ledgeway et al. (2022). https://doi.org/10.1093/oso/9780198870807.003.0010.
Finocchiaro, Chiara. 2005. Psychological evidence on the status of Romance clitics. *Italian Journal of Linguistics* 17 2: 291–310. Available at https://www.italian-journal-linguistics.com/2005-2/.
Finocchiaro, Chiara and Alfonso Caramazza. 2006. The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes* 21 1-3: 141–180. https://doi.org/10.1080/01690960400001887.
Firică, Camelia. 2009. The phonetic or the etymological principle in Romanian orthography? *Govor* 26 1: 53–62. Available at https://hrcak.srce.hr/165892.
Franco, Ludovico and Paolo Lorusso (eds.) . 2019. *Linguistic Variation: Structure and Interpretation*. De Gruyter Mouton. https://doi.org/10.1515/9781501505201.
Franks, S. and T.H. King. 2000. *A Handbook of Slavic Clitics*. Oxford Studies in Comparative Syntax. Oxford University Press. Available at https://global.oup.com/academic/product/a-handbook-of-slavic-clitics-9780195135886.
Gerlach, B. and J. Grijzenhout (eds.) . 2001. *Clitics in Phonology, Morphology and Syntax*. John Benjamins. https://doi.org/10.1075/la.36.
Gerlach, Birgit. 2002. *Clitics Between Syntax and Lexicon*. John Benjamins. https://doi.org/10.1075/la.51.
Gerstenberger, Ciprian-Virgil. 2007. A mereology-based general linearization model for surface realization. In *Proceedings of EUROLAN 2007*. University of Iaşi, Romania. Available at https://www.researchgate.net/publication/233951937_A_mereology-based_general_linearization_model_for_surface_realization.
Gerstenberger, Ciprian-Virgil. 2018. A grammar for romanian weak pronoun generation. In *Languages at the Crossroads: Training, Accreditation and Context of Use. Proceedings of the 35th Edition of the International Conference of the Spanish Association of Applied Linguistics*, edited by Francisco Javier Díaz Pérez and M.ª Águeda Moreno Moreno, pp. 215–226. University of Jaén. Available at https://www.researchgate.net/publication/332671041_A_Grammar_for_Romanian_weak_pronoun_generation.
Giurgea, Ion. 2013. L'origine des clitiques roumains de 3e personne pluriel datif et de 1ere personne pluriel datif-accusatif. *Revue roumaine de linguistique* LVIII 2: 131–143.
Graur, Al. 1960. Observaţii asupra sinerezei în romîneşte [Observations concerning synaeresis in Romanian]. *Studii şi Cercetări Lingvistice* .
Guţu-Romalo, Valeria (ed.) . 2008. *Gramatica limbii române*. Editura Academiei Române.
Harris, Alice C. 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford University Press. Available at https://global.oup.com/academic/product/endoclitics-and-the-origins-of-udi-morphosyntax-9780199246335.
Haspelmath, Martin. 2015. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In Blaszczak et al. (2015), pp. 273–304. https://doi.org/10.1515/9781614514510-009.
Haspelmath, Martin. 2022. Types of clitics in the world's languages. In preparation; available at https://www.academia.edu/s/52c1b938dc.
Hill, Virginia and Gabriela Alboiu. 2016. *Verb Movement and Clause Structure in Old Romanian*. Oxford Studies in Diachronic and Historical Linguistics. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780198736509.001.0001.
Iliescu, Maria. 1975. Pentru o sistematizare a predării pronumelui personal neaccentuat românesc (la studenţii străini). *Limba Română* 24: 51–62. Available at https://www.diacronia.ro/en/indexing/details/A11324.

Jianu, Maria-Magdalena. 2013. Observations upon the Clitics of the Dative Case in Romanian. *International Journal of Communication Research* 3 2: 117.

Johnson, Sally. 2005. *Spelling Trouble?: Language, Ideology and the Reform of German Orthography*. Multilingual Matters. Available at https://www.multilingual-matters.com/page/detail/Spelling-Trouble-Language-Ideology-and-the-Reform-of-German-Orthography/?k=9781853597848.

Kabatek, Johannes, Philipp Obrist, and Albert Wall (eds.) . 2021. *Differential Object Marking in Romance: The third wave*. De Gruyter. https://doi.org/10.1515/9783110716207.

Klavans, Judith L. (ed.) . 1995. *On Clitics and Cliticization: The Interaction of Morphology, Phonology, and Syntax*. Routledge. https://doi.org/10.4324/9780429442636.

Klein, Udo-Michael. 2007. *Encoding of argument structure in Romanian and SiSwati*. Ph.D. thesis, University of London. Available at http://hdl.handle.net/11858/00-001M-0000-0012-8EEC-D.

Krämer, Martin. 2012. *Underlying Representations*. Cambridge University Press. https://doi.org/10.1017/CBO9780511978821.

Ledgeway, Adam, John Charles Smith, and Nigel Vincent (eds.) . 2022. *Periphrasis and Inflexion in Diachrony: A View from Romance*. Oxford Studies in Diachronic and Historical Linguistics Series. Oxford University Press. https://doi.org/10.1093/oso/9780198870807.001.0001.

Legendre, Géraldine. 2001. Positioning Romanian verbal clitics at PF. In Gerlach and Grijzenhout (2001). https://doi.org/10.1075/la.36.10leg.

Lombard, Alf. 1976. Le î prosthétique du roumain. *Acta Societatis linguisticae Upsaliensis* 2 5.

Luraghi, Silvia. 2017. Clitics. In Luraghi and Parodi (2017), chap. 11, p. 165–193. https://doi.org/10.5040/9781472542090.

Luraghi, Silvia and Claudia Parodi (eds.) . 2017. *The Bloomsbury Companion to Syntax*. Continuum. https://doi.org/10.5040/9781472542090.

Maiden, Martin, Adina Dragomirescu, Gabriela Pană Dindelegan, Oana Uţă, and Rodica Zafiu. 2021. *The Oxford History of Romanian Morphology*. Oxford University Press. https://doi.org/10.1093/oso/9780198829485.001.0001.

Manzini, M. Rita. 2014. Grammatical categories: Strong and weak pronouns in Romance. *Lingua* 150: 171–201. https://doi.org/10.1016/j.lingua.2014.07.001.

Marioţeanu, Matilda Caragiu (ed.) . 1994. *Di nuntru şi-di nafoarâ: Stihuri armâneşti*. Cartea Românească.

Mišeska Tomić, Olga. 2006. *Balkan Sprachbund Morpho-Syntactic Features*. Springer. https://doi.org/10.1007/1-4020-4488-7.

Monachesi, Paola. 2001. Clitic placement in the Romanian verbal complex. In Gerlach and Grijzenhout (2001). https://doi.org/10.1075/la.36.11mon.

Monachesi, Paola. 2005. *The Verbal Complex in Romance: A Case Study in Grammatical Interfaces*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199274758.001.0001.

Montreuil, Jean-Pierre Y. (ed.) . 2006. *New Perspectives on Romance Linguistics*. John Benjamins. https://doi.org/10.1075/cilt.276.

Nastasenco, O. A. 1997. *Gramatica limbii române în tabele*. Editura VIRT, Chişinău.

Nevins, Andrew and Oana Săvescu. 2008. An Apparent 'Number Case Constraint' in Romanian: The Role of Syncretism. *Romance Linguistics 2008. Selected Papers from the 36th Linguistic Symposium on Romance Languages* https://doi.org/10.1075/cilt.313.17nev.

Ordóñez, Francisco and Lori Repetti. 2006. Stressed Enclitics? In Montreuil (2006). https://doi.org/10.1075/cilt.276.13ord.

Pescarini, Diego. 2018. Stressed enclitics are not weak pronouns: A plea for allomorphy. In *Romance Languages and Linguistic Theory 14: Selected papers from the 46th Linguistic Symposium on Romance Languages (LSRL)*. Stony Brook. https://doi.org/10.1075/rllt.14.13pes.

Pescarini, Diego. 2019. An emergentist view on functional classes. In Franco and Lorusso (2019), pp. 531–560. https://doi.org/10.1515/9781501505201-027.

Popescu, Alexandra. 2000. The morphophonology of the Romanian clitic sequence. In *Lingua*, vol. 110, pp. 773–799. Elsevier. https://doi.org/10.1016/S0024-3841(00)00016-4.

Popescu, Alexandra. 2003. *Morphophonologische Phänomene des Rumänischen*. Ph.D. thesis, University of Düsseldorf. Available at https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=3187.

van Riemsdijk, Henk C. 1999a. Clitics: A state-of-the-art report. In van Riemsdijk (1999b). https://doi.org/10.1515/9783110804010.1.

van Riemsdijk, Henk C. (ed.) . 1999b. *Clitics in the Languages of Europe*. Mouton de Gruyter. https://doi.org/10.1515/9783110804010.

Rosetti, Alexandru. 1986. *Istoria limbii române*. Editura ştiinţifică şi enciclopedică.

Rouveret, Alain. 1999. Clitics, subjects and tense in European Portuguese. In van Riemsdijk (1999b). https://doi.org/10.1515/9783110804010.

Sasaki, Kan and Daniela Căluianu. 2000. An Optimality Theoretic Account for the Distribution of Pronominal Clitics in Romanian. Tech. rep., University of Tsukuba. Available at https://www.academia.edu/33668351/An_optimality_theoretic_account_for_the_distribution_of_pronominal_clitics_in_Romanian.

Somesfalean, Stanca. 2007. *On the Form and Interpretation of Clitics*. Ph.D. thesis, Université du Québec à Montréal. Available at https://archipel.uqam.ca/9628/.

Spencer, Andrew and Ana Luís. 2012. *Clitics: An Introduction*. Cambridge University Press.

Săvescu Ciucivara, Oana. 2009. *A Syntactic Analysis of Pronominal Clitic Clusters in Romance - The view from Romanian*. Ph.D. thesis, New York University. Available at https://www.proquest.com/docview/304954033.

Sweller, John, Jeroen J. G. van Merriënboer, and Fred Paas. 2019. Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review* 31 2: 261–292. https://doi.org/10.1007/s10648-019-09465-5.

Tigău, Alina. 2021. Differential Object Marking in Romanian and Spanish: A contrastive analysis between differentially marked and unmarked direct objects. In Kabatek et al. (2021), pp. 173–212. https://doi.org/10.1515/9783110716207-007.

Zafiu, Rodica. 2019. Schimbarea morfologică în paradigma de prezent a verbului „a fi": formele bănăţene „mi(-)s", „ni(-)s", „vi(-)s". In Dindelegan et al. (2019). Available at https://editura-unibuc.ro/en/magazin/filologie/limba-romana/variatie-diacronica-si-diatopica-note-gramaticale/.

Zwicky, Arnold. 1977. On clitics. *Bloomington, IN: Indiana University Linguistics Club* Available at https://arnoldzwicky.org/about/papers/.

Zwicky, Arnold. 1985. Clitics and Particles. *Language* 61 2: 283–305. https://doi.org/10.2307/414146.

Zwicky, Arnold M. and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 59 3: 502–513. https://doi.org/10.2307/413900.

Appendix 1: Hyphen and syllabicity disambiguation for pre- and post-verbal Romanian weak pronouns

Example	**[aɪ pʰiː eɪ]**	***Orthography***
Gloss	**Disambiguation**	**'Translation'**
(3) *î -l cumperi* î$_{\text{host}}$ cl$_{\text{3.sg.acc.m}}$ buy$_{\text{2.sg.pres}}$	[ɨl.ˈcum.perⁱ] [î] (l) cumperi	*Îl cumperi.* 'You buy it.'
(4) *cumpără -l* buy$_{\text{2.sg.imp}}$ cl$_{\text{3.sg.acc.m}}$	[ˈcum.pə.rəl] [cumpără] —▬ (l)	*Cumpără-l!* 'Buy it!'
(5) *ți -l cumperi* cl$_{\text{2.sg.dat}}$ cl$_{\text{3.sg.acc.m}}$ buy$_{\text{2.sg.pres}}$	[tsil.ˈcum.perⁱ] [ți] ▬ (l) cumperi	*Ți-l cumperi.* 'You buy it.'
(6) *cumpără -ți -l* buy$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.dat}}$ cl$_{\text{3.sg.acc.m}}$	[ˈcum.pə.rə.tsil] cumpără —— [ți] —▬ (l)	*Cumpără-ți-l!* 'Buy it!'
(7) *ți- o cumperi* cl$_{\text{2.sg.dat}}$ cl$_{\text{3.sg.acc.f}}$ buy$_{\text{2.sg.pres}}$	[tsi̯o.ˈcum.perⁱ] (ți̯) ▬ [o] cumperi	*Ți-o cumperi.* 'You buy it.'
(8) *cumpără -ți -o* buy$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.dat}}$ cl$_{\text{3.sg.acc.f}}$	[ˈcum.pə.rə.tsi̯o] cumpără —— (ți̯) —▬ [o]	*Cumpără-ți-o!* 'Buy it!'
(9) *să -ți cumperi cartea* that cl$_{\text{2.sg.dat}}$ buy$_{\text{2.sg.pres}}$ book$_{\text{def}}$	[sətsⁱ.ˈcum.perⁱ.ˈcar.te̯a] [să] ▬ (ți̯) cumperi cartea	*Să-ți cumperi cartea!* 'Buy the book!'
(10) *cumpără -ți cartea* buy$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.dat}}$ book$_{\text{def}}$	[ˈcum.pə.rətsⁱ.ˈcar.te̯a] [cumpără] —▬ (ți̯) cartea	*Cumpără-ți cartea!* 'Buy the book!'
(11) *ți- ai cumpărat cartea* cl$_{\text{2.sg.dat}}$ have$_{\text{2.sg.pres}}$ buy$_{\text{past.part}}$ book$_{\text{def}}$	[tsi̯ai̯.cum.pə.ˈrat.ˈcar.te̯a] (ți̯) ▬ [ai] cumpărat cartea	*Ți-ai cumpărat cartea.* 'You've bought the book.'
(12) *te duci acasă* cl$_{\text{2.sg.acc.refl}}$ carry$_{\text{2.sg.pres}}$ home	[te.ˈdutʃⁱ.a.ˈca.sə] [te] duci acasă	*Te duci acasă.* 'You go home.'
(13) *du -te acasă* carry$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.acc.refl}}$ home	[ˈdu.te.a.ˈca.sə] du —— [te] acasă	*Du-te acasă!* 'Go home!'
(14) *du -te -acasă* carry$_{\text{2.sg.imp}}$ cl$_{\text{2.sg.acc.refl}}$ home	[ˈdu.te̯a.ˈca.sə] du —— (te) ▬ [acasă]	*Du-te-acasă!* 'Go home!'
(15) *mi le faci acum* cl$_{\text{1.sg.dat}}$ cl$_{\text{2.pl.acc.f}}$ do$_{\text{2.sg.pres}}$ now	[mi.le.ˈfatʃⁱ.a.ˈcum] [mi] [le] faci acum	*Mi le faci acum.* 'You do them for me now.'
(16) *fă -mi -le acum* do$_{\text{2.sg.imp}}$ cl$_{\text{1.sg.dat}}$ cl$_{\text{2.pl.acc.f}}$ now	[ˈfə.mi.le.a.ˈcum] fă —— [mi] —— [le] acum	*Fă-mi-le acum!* 'Do them for me now!'
(17) *fă -mi -le -acum* do$_{\text{2.sg.imp}}$ cl$_{\text{1.sg.dat}}$ cl$_{\text{2.pl.acc.f}}$ now	[ˈfə.mi.le̯a.ˈcum] fă —— [mi] —— (le) ▬ [acum]	*Fă-mi-le-acum!* 'Do them for me now!'

Legend: [xx] → **syllabic weak pronoun** (xx) → **asyllabic weak pronoun**
[xx] → **syllabic host** [[xx]] → **syllabic weak pronoun and syllabic host**
[xx] ▬ (yy) → **sandhi marker** xx —— [yy] → **postverbal marker** [xx] —▬ (yy) → **sandhi and postverbal marker**

Appendix 2: Examples of analyzed Romanian weak pronouns

Glossed examples	**Orthography**	**IPA**	**Translation**
(18) dă -l give$_{_2.sg.imp}$ cl$_{_3.sg.acc.m}$	*Dă-l!*	[ˈdəl]	‘Give it!’
(19) a -l da to cl$_{_3.sg.acc.m}$ give$_{_inf}$	*a-l da*	[al.ˈda]	‘to give it’
(20) a î -l da to î$_{_host}$ cl$_{_3.sg.acc.m}$ give$_{_inf}$	*a îl da*	[a.ɨl.ˈda]	‘to give it’
(21) dându -l give$_{_ger}$ cl$_{_3.sg.acc.m}$	*dându-l*	[ˈdɨn.dul]	‘giving it’
(22) nu -l da not cl$_{_3.sg.acc.m}$ give$_{_inf}$	*Nu-l da!*	[ˈnul.ˈda]	‘Don’t give it!’
(23) nu î -l da! not î$_{_host}$ cl$_{_3.sg.acc.m}$ give$_{_inf}$	*Nu îl da!*	[ˈnu.ɨl.ˈda]	‘Don’t give it!’
(24) mi -l dai cl$_{_1.sg.dat}$ cl$_{_3.sg.acc.m}$ give$_{_2.sg.pres}$	*Mi-l dai.*	[mil.ˈdaɪ̯]	‘You give it to me.’
(25) dă -le give$_{_2.sg.imp}$ cl$_{_3.pl.acc.f}$	*Dă-le!*	[ˈdə.le]	‘Give them!’
(26) mi le dai acum cl$_{_1.sg.dat}$ cl$_{_3.pl.acc.f}$ give$_{_2.sg.pres}$ now	*Mi le dai acum.*	[mi.le.ˈdaɪ̯.a.ˈcum]	‘You give them to me now.’
(27) dă -mi -le acum give$_{_2.sg.imp}$ cl$_{_1.sg.dat}$ cl$_{_3.pl.acc.f}$ now	*Dă-mi-le acum!*	[ˈdə.mi.le.a.ˈcum]	‘Give them to me now!’
(28) dă -mi -le -acum give$_{_2.sg.imp}$ cl$_{_1.sg.dat}$ cl$_{_3.pl.acc.f}$ now	*Dă-mi-le-acum!*	[ˈdə.mi.leˌa.ˈcum]	‘Give them to me now!’
(29) se -ntâmplă cl$_{_3.acc.refl}$ happen$_{_3.sg.pres}$	*Se-ntâmplă.*	[sen.ˈtɨm.plə]	‘It happens.’
(30) se întâmplă cl$_{_3.acc.refl}$ happen$_{_3.sg.pres}$	*Se întâmplă.*	[se.ɨn.ˈtɨm.plə]	‘It happens.’
(31) î -l pot vedea î$_{_host}$ cl$_{_3.sg.acc.m}$ can$_{_1.sg.pres}$ see$_{_inf}$	*Îl pot vedea.*	[ɨl.ˈpot.veˈdeˌa]	‘I can see him/it.’
(32) copiii mi -s acolo child$_{_pl.def}$ cl$_{_1.sg.dat}$ be$_{_3.pl.pres}$ there	*Copiii mi-s acolo.*	[co.ˈpi.i̯i.mis.a.ˈco.lo]	‘My children are there.’
(33) acolo mi -s copiii there cl$_{_1.sg.dat}$ be$_{_3.pl.pres}$ child$_{_pl.def}$	*Acolo mi-s copiii!*	[a.ˈco.lo.mis.co.ˈpi.i̯i]	‘There are my children!’
(34) mi -o dai cl$_{_1.sg.dat}$ cl$_{_3.sg.acc.f}$ give$_{_2.sg.pres}$	*Mi-o dai.*	[mi̯o.ˈdaɪ̯]	‘You give her/it to me.’
(35) mi -ai dat -o cl$_{_1.sg.dat}$ have$_{_2.sg.pres}$ given cl$_{_3.sg.acc.f}$	*Mi-ai dat-o.*	[mi̯aɪ̯.ˈda.to]	‘You have given her/it to me.’
(36) te -aș vedea sănătos cl$_{_2.sg.acc}$ have$_{_1.sg.cond}$ see$_{_inf}$ healthy	*Te-aș vedea sănătos!*	[teˌaʃ.ve.ˈdeˌa.sə.nə.ˈtos]	‘May I see you healthy!’
(37) vedea -te -aș sănătos see$_{_inf}$ cl$_{_2.sg.acc}$ have$_{_1.sg.cond}$ healthy	*Vedea-te-aș sănătos!*	[ve.ˈdeˌa.teˌaʃ.sə.nə.ˈtos]	‘May I see you healthy!’
(38) vedea -o -aș moartă see$_{_inf}$ cl$_{_3.sg.acc.f}$ have$_{_1.sg.cond}$ dead	*Vedea-o-aș moartă!*	[ve.ˈdeˌa.o̯aʃ.ˈmo̯ar.tə]	‘May I see her dead!’
(39) mi l- ai dat cl$_{_1.sg.dat}$ cl$_{_3.sg.acc.m}$ have$_{_2.sg.pres}$ given	*Mi l-ai dat.*	[mi.laɪ̯.ˈdat]	‘You have given it to me.’
(40) că -mi dai mere that cl$_{_1.sg.dat}$ give$_{_2.sg.pres}$ apples	*că-mi dai mere*	[cəmʲ.ˈdaɪ̯.ˈme.re]	‘that you give me apples’
(41) că î -mi dai mere that î$_{_host}$ cl$_{_1.sg.dat}$ give$_{_2.sg.pres}$ apples	*că îmi dai mere*	[cə.ɨmʲ.ˈdaɪ̯.ˈme.re]	‘that you give me apples’
(42) mi- aduci mere cl$_{_1.sg.dat}$ bring$_{_2.sg.pres}$ apples	*Mi-aduci mere.*	[mi̯a.ˈdutʃʲ.ˈme.re]	‘You bring me apples.’
(43) î -mi aduci mere. î$_{_host}$ cl$_{_1.sg.dat}$ bring$_{_2.sg.pres}$ apples	*Îmi aduci mere.*	[ɨmʲ.a.ˈdutʃʲ.ˈme.re]	‘You bring me apples.’
(44) roua mi -mbată inima dew$_{_def}$ cl$_{_1.sg.dat}$ makes drunk heart$_{_def}$	*Roua mi-mbată inima.*	[ˈro.wa.mim.ˈba.tə.ˈi.ni.ma]	‘The dew makes my heart drunk.’
(45) roua î -mi îmbată dew$_{_def}$ î$_{_host}$ cl$_{_1.sg.dat}$ makes drunk inima. heart$_{_def}$	*Roua îmi îmbată inima.*	[ˈro.wa.ɨmʲ.ɨm.ˈba.tə.ˈi.ni.ma]	‘The dew makes my heart drunk.’