Lauzon, Balthazar & Folli, Raffaella & Sevdali, Christina & Gerard, Juliana. 2025. Acquisition of auxiliary selection in French and Italian and the role of input. *Glossa: a journal of general linguistics* 10(1). pp. 1–42.
DOI: https://doi.org/10.16995/glossa.16417

Open Library of Humanities

Acquisition of auxiliary selection in French and Italian and the role of input

Balthazar Lauzon, Ulster University, 2-24 York Street, Belfast, BT15 1AP, UK, b.lauzon@ulster.ac.uk

Raffaella Folli, Ulster University, r.folli@ulster.ac.uk

Christina Sevdali, Ulster University, c.sevdali@ulster.ac.uk

Juliana Gerard, Ulster University, j.gerard@ulster.ac.uk

This paper focuses on child acquisition of perfective auxiliaries in French and Italian by analysing auxiliary productions in naturalistic corpora and considering the role of input. Unlike languages like English or Spanish, which use *have* with all verb types, French and Italian use both *be* and *have* to express the perfective. *Be*-selection is a general marker for unaccusative and reflexive verbs, while unergative and transitive verbs select *have*. Previous work has shown that there exists considerable variation in auxiliary selection in some varieties of French, with increased use of *have* in contexts which require *be* in standard French. In acquisition studies, it is not yet clear whether non-adultlike child productions in French are evenly distributed or restricted to *be*-selecting verbs, while an asymmetry by person/number form has been reported for reflexives. Our results show firstly that child errors in French are indeed restricted to *be*-selecting verb types, but this pattern is specific to French and does not occur in child Italian. Importantly, the absence of variability in the input shows that these are child innovations. We further show that the asymmetry by person/number form is specific to 1SG *be* and occurs with intransitives in addition to reflexives. A more detailed analysis considering homophonous forms in spoken French demonstrates that these 1SG errors are not auxiliary selection errors but instead represent extensions of 3SG *be* to 1SG *be* contexts due to the later production of the 1SG form of the *be* paradigm.

1 Introduction

1.1 Cross-linguistic differences in auxiliary selection

Some languages make use of two auxiliaries with intransitive verbs to express past perfective: *be* and *have*, as in (1). Auxiliary selection refers to the language-specific processes governing the choice of auxiliary. Languages that have this split include Germanic languages such as Dutch and German, Romance languages such as French and Italian as well as minority languages such as Occitan, Sardinian and Corsican. Other languages such as English and Spanish only have one perfective auxiliary (*have*), as in (2)[1] (Zaenen 1993; Hoekstra 1999; Legendre 2007; McFadden 2007; Rosemeyer 2014; Gillmann 2015; Kailuweit 2015).

(1) *French*
   a. Jean **est** parti.
      John be.3SG left
      'John left.'
   b. Jean **a** travaillé.
      Jean have.3SG worked
      'John worked.'

(2) *Spanish*
   a. Juan **ha** salido.
      John have.3SG left
      'John has left.'
   b. Juan **ha** trabajado.
      John have.3SG worked
      'John has worked.'

Intransitive verbs can be classified as unaccusatives, as in (1a, 2a), syntactically having a single internal argument, and unergatives, as in (1b, 2b), having a single external argument, as originally proposed by Perlmutter (1978) and Burzio (1986). The single (internal) argument of unaccusatives is initially assumed to take the same position as the object of transitive verbs, while the single (external) argument of unergatives is in the same initial position as the subject of transitives. With intransitives, the selection of *be* is a marker of unaccusativity. Unergatives (1b, 2b) and transitive verbs (3) select *have*.

(3) *French*
   Jean **a** vu Marie.
   Jean have.3SG seen Marie
   'Jean saw Marie'

[1] In French and Italian this present auxiliary–past participle combination can take either a perfect (anterior) or perfective interpretation. In contrast in English only a perfect interpretation is possible.

In addition to intransitive verbs, reflexive (pronominal) verbs in French (4) and Italian (5) select *be* (Manzini 1986; Zribi-Hertz 1987).

(4) *French*
Marie **s'est** lavée.
Marie REFL.3SG-be.3SG washed
'Marie washed (herself).'

(5) *Italian*
Maria **si** **è** lavata.
Maria REFL.3SG be.3SG washed
'Maria washed (herself).'

There are important cross-linguistic differences in auxiliary selection rules. Languages differ in the number of intransitive verbs that select *be*. French *be*-selecting intransitives are considered a subset of Italian unaccusatives (Sorace 2000; Legendre 2007), with approximately 90 in Italian and 20 in French (Schwarze 1996; Heidinger 2015). However, this generalization may obscure other patterns. For instance, 'appear', 'happen' and 'be' select *be* in Italian, but form a mixed group in German, Dutch and French (Aranovich 2007). Specifically, the equivalent of unaccusative 'happen' selects *be* in both French (*arriver*) and Italian (*succedere*), while some stative verbs such as 'be' select *have* in French, but *be* in Italian. Also, semantically similar verbs may not select the same auxiliary, such as 'disappear' and 'appear', which select different auxiliaries in French (*disparaître* 'disappear' selects *have* while *apparaître* 'appear' typically selects *be*). Both select *be* in Italian (*scomparire/aparire*).

In contrast to intransitives, reflexive verbs show limited cross-linguistic differences. As mentioned above, these verbs all select *be* in both Italian and French. Semantic reflexives in German and Dutch, which take a reflexive pronoun such as 3SG 'zich' (Dutch) / 'sich' (German) select *have* similarly to transitive verbs, as in (6) (McFadden 2007: 691).

(6) *German*
Sepp **hat** sich verletzt.
Sepp have.3SG himself hurt
'Sepp hurt himself.'

Table 1 summarizes these auxiliary selection rules.

Theoretical accounts generally differ on whether auxiliary selection is governed by lexical semantics or syntax. For lexical semantics-based proposals, determining factors include categorisations based on change of state, stativity or telicity (Bentley & Eythórsson 2004; Randall 2007). For instance, Bentley and Eythórsson (2004) propose that *have* is a default (elsewhere)

	unaccusative	**unergative**	**reflexive**	**transitive**
	internal argument	external argument	reflexive pronoun	two arguments
French/Italian	BE	HAVE	BE	HAVE
Spanish/English	HAVE	HAVE	HAVE	HAVE
French example	Jean est parti.	Jean a travaillé.	Jean s'est lavé.	Jean a vu M.

Table 1: Summary of auxiliary selection rules.

auxiliary while *be* is produced as a morphological marker when semantic conditions are met. An influential framework for intransitive verbs is the Auxiliary Selection Hierarchy (ASH), first proposed by Sorace (Sorace 2000; Legendre & Sorace 2010), which divides intransitives into separate groups along agentive-aspectual lines going from categorical *be*-selecting (change of location, e.g. 'arrive') to categorical *have*-selecting (controlled processes, e.g. 'work') verbs, one key aim being to account for cross-linguistic differences such as in the case of 'existence of state' verbs (e.g. 'exist' which takes *be* in Italian and *have* in French).

Alternatively, syntax-based accounts argue that auxiliary selection with intransitives is not determined by the projection of argument structure (internal or external argument) but by the specific syntactic frame the verb is inserted in (Borer 1994; Folli 2001), which may better account for 'variable behaviour' verbs such as *correre* (run), as in (7) (van Valin 1990: 233).

(7) *Italian*

a. Luisa **ha** corso nel parco (per un ora).
Luisa have.3SG run in-the park (for an hour)
'Luisa ran in the park (for an hour).'

b. Luisa **è** corsa a casa (in un ora).
Luisa is.3SG run to home (in an hour)
'Luisa ran home (in an hour).'

Variability thus appears to the extent that these verbs occur in different syntactic frames. Other syntax-based proposals view auxiliary selection as the product of a local Agree (probe-goal) relation with a functional head (Bjorkman 2011; Amato 2023). According to Amato's (2023) proposal, whether auxiliary *be* or *have* is spelled out depends on specific features of the functional head Perf. The place of reflexives is also a matter of debate within syntactic approaches, though most authors view the reflexive clitic as a valency reducing element rather than a true pronoun (see for instance McGinnis 2022). Some accounts analyse Romance reflexives as essentially unaccusative, with a single internal argument and reduction of the external argument (Grimshaw 1992; Embick 2004), therefore accounting for the selection of *be*. However, others argue that the external argument is preserved and it is the internal argument that is reduced, at

least for agent arguments such as in (4) and (5) above. *Be*-selection in this case is proposed to be the result of either the missing theta-role or the missing accusative case assignment (Reinhart & Siloni 2005).

In the following section, we show that there is additional intra-dialectal variation that must be taken into account by studies of auxiliary selection.

1.2 Synchronic variation

Varieties of French that have been shown to have greater use of *have* with *be*-selecting intransitive and reflexive verbs include North American regions: Montreal French, Vermont French, Acadian French as well as other French minority regions in Canada (Sankoff & Thibault 1977; Russo & Roberts 1999; Willis 2000; King & Nadasdi 2005; Rea 2020).

Studies of variation in Montreal French using corpora generated from recorded participant interviews have found significant rates of *have* with *be*-selecting intransitive verbs (Sankoff & Thibault 1977; Rea 2020), ranging from less than 10% (e.g. *aller* 'go', *venir* 'come', *partir* 'leave') to 90% *have* productions (e.g. *tomber* 'fall', *passer* 'pass').[2] There was some degree of variation for all participants in these studies, although the speaker's level of education correlated with a generally lower rate of *have* use. Item frequency was inversely correlated with *have* use, i.e. *be* was more likely with higher frequency items. These studies also show that variation is not evenly distributed: some speakers (almost) never produce non-standard auxiliaries, while others do so frequently, and intra-speaker variability is observed for some speakers. This was true for both intransitive and reflexive verbs. For instance, in Rea (2020) non-standard *have* use with reflexives was at a low frequency (1.3% of total) and restricted to a minority of participants, with individual rates ranging from approximately 2% to 38%.

Other North American varieties show even greater variability. Acadian French (Eastern Canada) has been found to have near-categorical use of *have* with intransitive and reflexive verbs in some of the French-speaking regions studied (King & Nadasdi 2005; Balcom 2008). King and Nadasdi (2005) for instance find that in one region the only verbs displaying some variability are *mourir* 'die' and *naître* 'be born', where auxiliary use can be linked to the contrast between eventive and resultative uses, as in (8) (King & Nadasdi 2005: 106).[3]

(8) *Acadian French*

Il	**est**	**mort**	asteure.	Il	**a**	**mouri**	avant
he	be.3SG	died	now	he	have.3SG	died	before.

'He is dead now. He died before.'

[2] *Have* frequency by verb in these two studies is similar though not identical. For instance, the percentage *avoir* use for *partir* 'leave' is relatively high in Sankoff & Thibault (1977) but low in Rea (2020). See **Table 3** for the list of *be*-selecting intransitives analysed in the present study.

[3] For this verb, *have* is used with the non-standard participle 'mouru/mouri'.

A study of Vermont French (Russo & Roberts 1999) showed similar results overall to the Montreal study by Sankoff and Thibault (1977), with significant differences in auxiliary use by verb.

One possible factor for the increased rate of *have* production with *be*-selecting verbs is contact with English as these varieties of French occur in areas that are likely to have a high level of bilingualism. This is supported by studies that found *have* use to be correlated with a lower French proficiency in English-French bilinguals (Canale et al. 1978; Mougeon & Beniak 1986). It is therefore possible that non-standard *have* auxiliary uses are in part evidence for change through contact with English. If so, European varieties may show more limited variation due to generally lower levels of bilingualism. However, it is also possible that the dominant factors responsible for this variability are socio-economic, and in this case we could expect to find similar variability in European varieties of French.

Relatively less is known about variation in Italian. The Auxiliary Selection Hierarchy makes explicit the prediction that societal variation is likely to be observed in the variable classes of verbs, such as 'existence of state' and 'continuation of a preexisting state', but not for categorical verb classes. In support of this, some experimental studies based on acceptability judgments (Bard et al. 1996) or eye-tracking (Vernice & Sorace 2018) are consistent with the view that speakers' preferences for either auxiliary is stronger for categorical than variable verb groups. However, Rastelli (2023) points out that corpus data do not necessarily match these predictions. For instance, *have* use with unaccusative past participles *durato* 'last', *rimasto* 'remain', *bastato* 'suffice', *appartenuto* 'belong' and *intervenuto* 'intervene' is negligible (Rastelli 2023: 73–74). An experimental study with university students in Northern Italy (Rastelli 2023) found that participants' choice of auxiliary could be considered 'variable' for only a subset of the expected verbs (approximately 60% of verbs considered variable based on the ASH). Additionally, participants' auxiliary preferences were generally better predicted by item frequency in the corpus than semantics (including the effect of telicity or classification into separate ASH categories for states and processes).

In summary, specific varieties are likely to present more variability in auxiliary production than what has been described for standard French and Italian. Despite considerable differences between speakers, factors associated with an increased rate of *have* auxiliaries include: contact with English, socio-economic factors representing a lower awareness of normative usage, and lower item frequency. This shows that variability in participants' auxiliary use must be taken into consideration. For studies of acquisition in particular this underscores the need to consider potential input variability as children's productions (or comprehension) would be different in the presence of variable linguistic input.

1.3 Acquisition of auxiliary selection

Prior studies on the acquisition of auxiliary selection in French and Italian have analysed child error[4] rates in production of auxiliaries with different verb types but reach different conclusions (Snyder et al. 1995; Boyce et al. 2017). Firstly, their results differ on whether child errors occur with both auxiliaries or preferentially represent the overuse of *have* with *be*-selecting verbs. Secondly, the studies disagree on the overall rate of error and whether the same error patterns are seen with reflexives and intransitives. Furthermore, the studies did not include an analysis of language input which, as argued above, is crucial for this investigation.

1.3.1 Child error patterns

Studying one French-speaking child and three Italian-speaking children from CHILDES corpora (MacWhinney 2000), Snyder et al. (1995) find low rates of error with *be*-selecting verb types (intransitive and reflexive), but no errors with *have*-selecting intransitives. In contrast, Boyce et al. (2017), reporting on a larger sample of French-speaking children from CHILDES, find that children make auxiliary selection errors with both *be*- and *have*-selection at a similar rate, i.e. equivalent frequencies of productions of *have* with *be*-selecting and *be* with *have*-selecting intransitives. Both studies find no auxiliary selection errors with transitive verbs.

Importantly, French reflexive and object clitics are homophonous for 1SG and 2SG: (9a) is a non-adultlike use of *have* with a 1SG reflexive clitic, while (9b) is an adultlike use of *have* with a 1SG object clitic.

(9) *French*

a. *je **m'ai** mordu
I REFL.1SG-have.1SG bitten
'I bit myself.'

b. il **m'a** mordu
he ACC.1SG-have.3SG bitten
'He bit me.'

Snyder et al. (1995) argue that children possess early representations of reflexive structures, due to the low rate of error (2 instances of *have* in 29 reflexive auxiliaries, and no occurrences of *be* in over 100 object clitic constructions). In contrast, Boyce et al. (2017) find an asymmetry in *have* use for 1SG, 2SG and 3SG, with the highest error rate for 1SG, no errors with 3SG and an intermediate rate for 2SG reflexives. They suggest that these differences could be due to the syncretism between reflexive and object clitics for 1SG and 2SG as in (9), but not 3SG clitics, as in (10).

[4] We use the term 'error' in this paper as a synonym for 'non-adultlike production', i.e. a child production that is different to what is expected for adult grammars, and without any connotation of inaccuracy or lack of ability.

(10) *French*

a. il **s'est** frappé
he REFL.3SG-be.3SG hit
'He hit himself.'

b. je **l'ai** frappé
I ACC.3SG-have.1SG hit
'I hit him/her.'

The argument is that productions of *have* with 1SG and 2SG reflexives are caused by the child's incorrect representation of transitive rather than reflexive syntax.

1.3.2 Remaining questions

The potential role of variation in language input was not measured in the studies discussed here. This is important as the presence of variability in the linguistic input would necessarily lead to a very different understanding of the nature of child errors, and since variability has been shown to occur, at least for some varieties (section 1.2). More specifically, if (non-standard) *have* uses are found in the language input, this means that children are likely to be reproducing this variability in their speech, whereas if we see no or very limited input variability this means that the errors are child innovations representing the incomplete acquisition of auxiliary selection.

Furthermore, these studies differ in the overall 'direction' of child errors with intransitives, as Snyder et al. (1995) find exclusively *have* use with *be*-selecting verbs, while Boyce et al. (2017) report equal rates for both error types. However, these two patterns have different implications for how auxiliary selection is acquired, namely whether *have* is first acquired by the child followed by a later acquisition of *be*-selection, or with both auxiliaries produced from early on.

Finally, it is not clear whether the same patterns of auxiliary selection errors are seen with intransitives and reflexives. Snyder et al. (1995) argue that *be*-selection in each of these represents the same underlying process: based on the unaccusative analysis of reflexives, syntactic knowledge of A-movement is needed for production of *be* with both unaccusatives and reflexives, and *have* with object clitic transitive constructions. On the other hand Boyce et al. (2017) argue for two different mechanisms based on the observed asymmetry of *have* frequencies by person/number form for reflexives, but not intransitives. Understanding exactly how auxiliary selection is acquired is important as it may lead to further insights on how this knowledge is represented.

1.4 Aims of the present study

The current study aims to address the conflicting findings that prior research has reported. This means analysing rates of error with *be*- and *have*-selecting intransitives in order to better understand how auxiliary selection is acquired. As detailed above, if errors are only observed for

be-intransitives, this would suggest that children acquire *have*-selection before *be*-selection. On the other hand, if both error patterns are observed (as in Boyce et al. 2017), this would suggest that both *be*- and *have*-selection are acquired simultaneously.

Secondly, we compare the acquisition of auxiliaries in French and Italian. Specifically, we aim to determine if overall error frequencies are similar in both languages. This is relevant as we could expect the same basic processes to determine the acquisition of auxiliary selection in Italian. If there are significant differences, we should ask whether there are potential causes that are independent of auxiliary selection.

We also analyse the role of input in auxiliary selection to determine if there is any variation. If there is some degree of non-standard use in parental speech, we would have to consider whether child errors are merely the result of reproducing this variable input and not due to grammar-internal factors.

Finally, by comparing child error patterns with intransitives and reflexives we can determine whether the acquisition of *be*-selection involves the same processes with both verb types. We therefore consider errors in relation to person/number forms for both reflexives and intransitives to determine if the reported person/number asymmetry (Boyce et al. 2017) is limited to reflexive constructions. If this also occurs with *be*-selecting intransitives, we could rule out the syncretism of clitic forms as a cause for such errors and look for other potential factors.

2 Methods

This study used naturalistic corpora to study auxiliary productions in child and adult speakers. The following sections detail the selection of the corpora, the search procedure for auxiliary–participle pairs, and coding.

2.1 Selection of corpora

This study used monolingual French corpora from CHILDES (MacWhinney 2000). We selected all the available corpora that spanned at least the period of approximately age 2 to 3 years, when auxiliaries are first produced. Corpora were also chosen for the availability of parental speech. The selected French corpora (and children) were: Champaud (Grégoire) (Champaud 2004), Geneva (Marie) (Hamann et al. 2003), Leveillé (Philippe) (Suppes et al. 1973), Lyon (Anaïs, Marie, Marilyn, Nathan, Théotime) (Demuth & Tremblay 2008), Paris (Anaé, Antoine, Julie, Léonard, Madeleine, Théophile) (Morgenstern & Parisse 2007), Yamaguchi (Adrien) (Yamaguchi 2012) and York (Anne, Léa and Max) (Plunkett 2002) (**Table 2**).

Most of the French-speaking children are from France, with the exception of Léa (Belgium), Max (Canada) and G-Marie (Geneva, Switzerland).

corpus	N	age	files	child utterances	parental utterances
Champaud	1	1;09–2;05	32	3,522	2,514
Geneva	1	1;08–2;06	17	5,951	6,459
Leveillé	1	2;01–3;03	33	14,430	13,747
Lyon	5	1–3	396	72,442	123,639
Paris	6	1–4	180	48,751	86,046
Yamaguchi	1	1;03–4	31	7,998	12,959
York	3	1;10–3;05	107	26,368	10,077
Total	**18**	–	**796**	**179,462**	**255,441**

Table 2: Selected French corpora.

2.2 Searches, coding procedures and exclusions

Searches were conducted using the CLAN program (MacWhinney 2000). Different search procedures were used depending on verb type (*be*-selecting intransitive, *have*-selecting intransitive, reflexive).[5] The first series of searches were conducted on the Leveillé corpus and aimed to replicate the number of *be* and *have* auxiliaries with reflexive verbs from Snyder et al. (1995).

2.2.1 French searches

Generally, initial searches were carried out for auxiliary/past participle or clitic-auxiliary pairs, followed subsequently by more specific searches, though separate search terms were needed for the different verb types.

Based on the general searches, 19 separate French *be*-selecting intransitive verbs were found in auxiliary–participle constructions. We therefore ran a series of searches for each of these specific past participles (**Table 3**).[6]

For French reflexive verbs, specific searches were based on the presence of a clitic and auxiliary. We could not search for each individual verb due to the large number of French reflexives.

[5] As shown in **Table 1**, the theory classifies verbs that have an internal (patient) argument and select *be* in the relevant languages as unaccusative, while verbs that take an agent external argument and select *have* as unergative. Here we use these category labels, following an anonymous reviewer's suggestion, to refer unambiguously to auxiliary selection rules, as unaccusativity is a larger issue than auxiliary selection. See for instance Aranovich (2007) and Heidinger (2015)

[6] French *be*-selecting intransitives were initially categorized as 'variable' or 'categorical' in order to distinguish verbs where *have* auxiliary use is possible, as it was not known whether variability would be observed in parental speech, with the categories 'variable' and 'categorical' based on frequencies of auxiliary *have* from Rea (2020) and Sankoff & Thibault (1977) and other studies of variation in auxiliary selection (section 1.2).

French participle	English	+/– transitive (exclusion rate)
tombé	fallen	–
allé	gone	–
parti	left	–
arrivé	arrived	–
passé	passed	+ (16%)
venu	come	–
monté	gone up	+ (15%)
rentré	gone in	+ (4%)
resté	stayed	–
sorti	gone out	+ (50%)
revenu	returned	–
devenu	become	–
descendu	gone down	+ (18%)
mort	died	–
reparti	left	–
né	been born	–
retourné	returned	+ (55%)
entré	entered	–
rendu	arrived	+

Table 3: French *be*-selecting past participles ordered by frequency.

Have-selecting intransitive verbs were obtained from the general searches for auxiliary–participle pairs. Note that this category includes true intransitives, such as *dormir* 'sleep', intransitives which can also be used transitively, such as *manger* 'eat', as well as some verbs that select *be* in Italian for instance, such as *durer* (last), since all of these select *have* in French.

For the search for the first occurrence of *be* as copula and auxiliary (section 3.4), we used separate keyword searches for each form (1SG and 3SG) of *be* and *have*, and searched the output for the earliest adultlike form in each case. Recordings and phonological transcription (where available) were used to confirm that the production of 1SG *be* 'suis' was adultlike.

See the supplementary files at https://doi.org/10.17605/OSF.IO/XWYGZ for further details on searches (including lists of CLAN commands, search procedures and verbs by category).

2.2.2 Ambiguous auxiliaries and exclusions

Utterances made up of songs and reading were excluded for both child and parental speech. Child repetitions of adult speech were excluded if they were identical to the parent's utterance,[7] but not if they were partial repetitions e.g. if a different person/number form was produced. Repetitions by the same speaker were generally counted as separate utterances if they were on separate lines in the transcript.

Utterances with auxiliary *be* and a participle were excluded if they were resultative uses, or if they were ambiguous. This applied mainly to *be*-selecting change of state intransitives, such as 'die' and *have*-selecting intransitives such as *casser* (French 'break').[8] For instance, (11a) (Léonard/Mother) was excluded as it is referring to an object, while (11b) (Léa, 3;05) was excluded as it is referring to the current state of a tape recording. However, 'stop' (11c) (Léa, 3;09) was counted as it refers to a repeated action.

(11) a. Il **est** déjà **sorti** il est dans la salle-à-manger.
it be.3SG already gone.out it be.3SG in the dining-room
'it is already out, it is in the dining room.'

b. Oui c'**est** déjà **arrêté**.
yes DEM.3SG-be.3SG already stopped
'Yes, it is already stopped.'

c. Alors il n'**a** pas **arrêté** de faire ça.
so he NEG-have.3SG NEG stopped of do that
'So he hasn't stopped doing that.'

In order to distinguish between anterior and resultative uses, we used the English translation as a criteria, and excluded any utterances that had English *be* as copula (with participle/adjective), such as (11a–b) above.

For intransitives that can be used transitively, such as *be*-selecting *monter* 'go up' or *have*-selecting *manger* 'eat', these were counted as long as there was no object or object clitic. Note that verbs with a single post-verbal argument can be ambiguous as post-verbal subjects are very common. In such cases, context (from transcripts and/or recordings) was used in order to

[7] An example of a direct repetition which was excluded is the following (Theotime, 2;03):

(i) MOT: éh y a l(e) capot qui s'est ouvert ('Hey, the bonnet just opened itself up')
CHI: le capot qui[?] s'est ouvert

[8] The exclusion rate for resultative use varied by verb. It was only significant for three of the *be*-selecting intransitives: *sortir* 'go out' (9% of occurrences), *monter* 'go up' (5% of occurrences) and *mourir* 'die' (61% of occurrences). For some *have*-selecting intransitives such as *arrêter* 'stop' and *casser* 'break', these uses made up the majority of occurrences (not counting reflexives *s'arrêter* and *se casser*) and for these verbs all occurrences with *be* were excluded.

determine the status of a post-verbal argument (whether subject or object). Total exclusion rates for transitive uses of the *be*-selecting intransitives are shown in **Table 3**.[9]

Reflexive verbs were excluded if no reflexive clitic was produced.

2.2.3 Italian corpora and searches

A similar general procedure was used to carry out searches for Italian. Similar criteria were used to select corpora and invidual children. The selected Italian corpora were: Antelmi (Cam) (Antelmi 1998), Tonelli (Elisa, Gregorio, Marco) (Tonelli et al. 1998) and Calambrone (Diana, Guglielmo, Martina, Raffaello, Rosa, Viola) (Cipriani 2004), for a total of 10 children (**Table 4**).

corpus	N	age	files	child utterances	parental utterances
Antelmi	1	2;02–3;04	7	1,835	1,275
Calambrone	6	1;08–2;10	79	19,259	21,487
Tonelli	3	1;05–2;05	43	10,784	16,929
Total	**10**	**–**	**129**	**31,878**	**39,691**

Table 4: Selected Italian corpora.

We needed fewer separate search commands for Italian due to the smaller size of the corpora compared to French. We used a general search for combinations of auxiliary (coded as 'v') and participle, with a few additional keyword searches for some specific auxiliary forms and some of the most common *be*-selecting intransitives such as *andato/a* 'gone'.

2.3 Coding and data analysis

Each selected utterance was coded for the individual child, language, speaker (child/adult), verb category (reflexive, *be*-selecting intransitive and *have*-selecting intransitive), specific verb, tense of the auxiliary (present or other), person/number (1SG, 2SG, 3SG, 1PL, 2PL, 3PL), the auxiliary produced in the utterance (*be* or *have*), and the selected auxiliary (according to language-specific auxiliary selection criteria). Each utterance was coded as 'adultlike' if the auxiliary produced matched the auxiliary selected, or 'non-adultlike' if there was a mismatch. Non-adultlike auxiliary uses were checked against the phonological tier (%pho) if available and/or the recording, where available.[10]

Data analysis was carried out in R (R Core Team 2023).

[9] *Rendu* is only equivalent to 'arrive' in the case of Max (York corpus) and is either transitive or reflexive for all other individuals.

[10] For utterances not containing a pre-verbal subject, the context (based on transcripts and recordings) was sometimes needed in order to distinguish between present tense 1SG *have* ('ai') and 3rd-person singular *be* ('est'), as these are homophonous in spoken French.

3 Results

3.1 Auxiliary use in French and Italian input

As discussed above, we aim to determine if children's non-adultlike auxiliary uses in French are true errors or can be attributed to variability in the input. As a first step, we therefore analyse auxiliaries produced in parental speech to determine the degree of variability that is present, if any, for each verb type: reflexive, *be*-selecting intransitive and *have*-selecting intransitive.

If child errors are the result of input variability, we can expect to see non-standard uses in both child and adult data. On the other hand, if children's non-adultlike auxiliary uses are innovations we would expect to find greater rates of such auxiliary productions in child speech than in the associated parental speech. Variability in this context is represented by a significant percentage of non-standard *have* auxiliaries with *be*-selecting intransitive and reflexive verbs as we would not expect any *be* auxiliaries with *have*-selecting intransitives in adult speech.

Auxiliaries in parental speech are shown in **Figure 1** as total counts of *be* and *have* auxiliaries by verb type for French (left) and Italian (right).

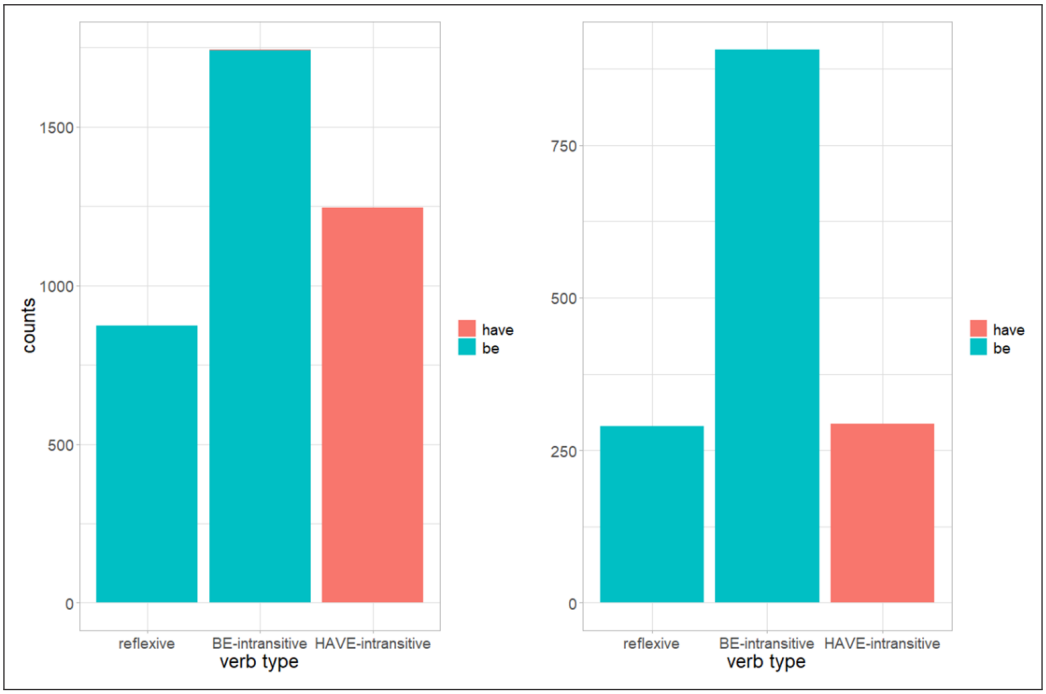

Figure 1: Auxiliary productions in French (left) and Italian (right) parental speech.

As this figure shows, in French there are no non-standard uses of *have* with reflexives (out of a total of 873 auxiliary–participle pairs) or of *be* with *have*-intransitives (out of a total of

1246 pairs). There are 3 counts of *have* with *be*-selecting intransitives (out of a total of 1741 auxiliary–participle pairs). These utterances are shown in (12).

(12) a. Il **a** **descendu**?
it have.3SG gone.down
'He went down?'

b. Je le tenais à peine et on **a** **descendu** comme ça et
I ACC.3SG hold.PST … hardly and we have.3SG gone.down like that and
il a pris un vrai téléski hein?
he have.3SG taken a real t-bar-lift eh
'I was barely holding on to him and we went down like that and he took a real t-bar lift'

c. Il a l'air bien hein quand même il **a** pas trop **monté** là
he have.3SG look good eh when even he have.3SG not too gone.up there
en température
in temperature
'He seems well enough and his temperature hasn't gone up too much'

In the case of (12a) the transcript reveals that the mother is repeating the child's utterance and auxiliary choice. For (12b) the verb *descendre* 'go down' is used transitively in a previous utterance (go down the slope) while for (12c) this is referring to the child's temperature rather than physical movement, usage which typically selects *have* (Manente 2008).

It is clear from these results that there is no real variability in the French language input. This means that any child errors must be child innovations as they cannot be the result of non-standard uses of *have* in the input.

For Italian (**Figure 1**, right), *be* is used with all *be*-selecting intransitive and reflexive verbs, while *have* is produced with all *have*-selecting intransitive verbs. This confirms that there is also no variability in the Italian input: as seen for French, parents produce auxiliaries according to auxiliary selection rules of standard Italian.

In the following section, we provide an analysis of auxiliaries produced by children, comparing productions of *be* and *have* for reflexive, *be*-selecting intransitive and *have*-selecting intransitive verbs.

3.2 Auxiliary use in child speech

We analyse the auxiliaries produced with intransitive and reflexive verbs in child data. Non-standard uses include the production of *have* with *be*-selecting intransitives and reflexive verbs, or that of *be* with *have*-selecting intransitives. Recall that previous studies (section 1.3.1) differed on whether children make errors with both auxiliaries or only overuse *have* with *be*-selecting

verbs. We therefore aim to determine if child errors are dependent on verb type. This is shown in **Figure 2** as total counts of *be* and *have* auxiliaries by verb type for French (left) and Italian (right).

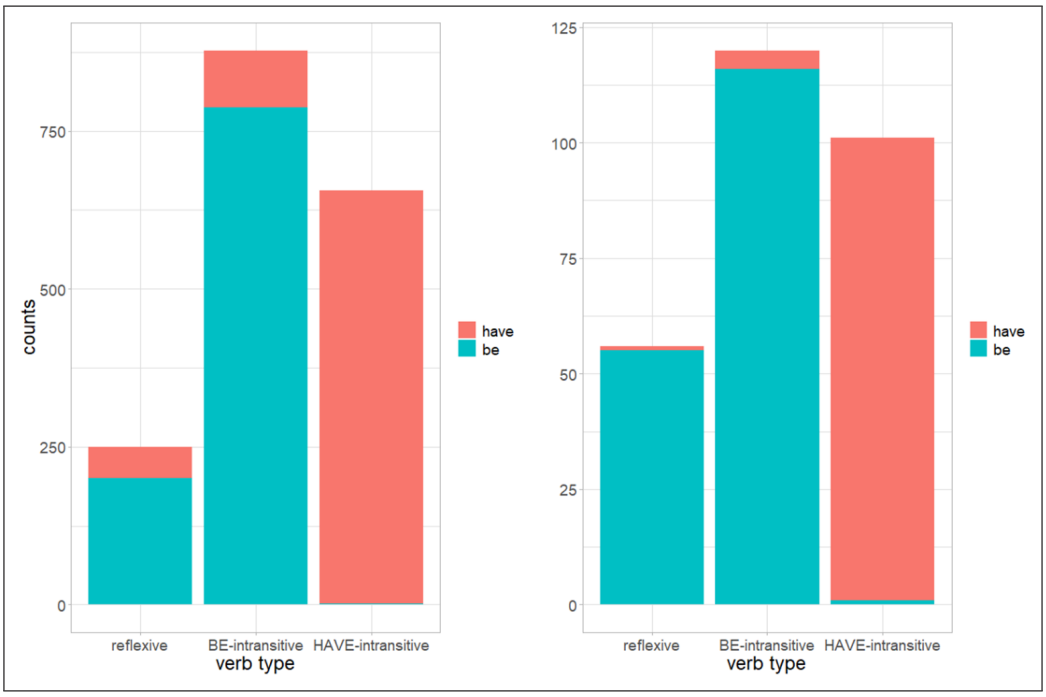

Figure 2: Auxiliary productions in child French (left) and Italian (right).

In French, the error frequencies are much higher for both *be*-selecting verb types, at 10% of auxiliaries with *be*-selecting intransitives (90 out of 878 counts) and 20% of auxiliaries with reflexives (50 out of 250 counts), than with *have*-selecting intransitives, where the frequency is less than 1% (2 out of 656 counts).

Examples of non-adult-like auxiliary uses are given below for French *be*-selecting intransitive (13a) (Anaé 2;06) and (13b) (Anaïs 2;09), and reflexive verbs (13c) (Grégoire 2;05).

(13) a. Moi j'**[e]** [ze] **allé** au bureau
me I-have.1SG/be.3SG … gone to.the office
'I went to the office.'

b. Oh oh oh il **a** **passé** d'un coup ici
oh oh oh he have.3SG passed of-a shot here
'Oh, he passed all of a sudden here'

c. Je **m'[e]** **cogné** ici
I REFL-have.1SG/be.3SG hit here
'I hit myself here'

To determine whether this pattern in French represents a significant deviation from the input, we ran a mixed effects logistic regression using R Statistical Software (R Core Team 2023) and the lme4 package (Bates et al. 2015) with production of auxiliary *have* as a binary dependent variable i.e. production of *have* was coded as 1, and production of *be* was coded as 0. Fixed effects included 'verb type' (reflexive, *be*-selecting and *have*-selecting intransitives) and 'speaker' (child/adult) with the individual child as a random intercept.

As the dependent variable is the production of *have*, we would expect to see a difference between *have*-selecting intransitives and reflexive or *be*-selecting intransitives as a main effect. More importantly, an interaction between 'speaker' and 'verb type' is indicative of a difference in *have* use across the different verb types between children and adults. If so, post-hoc tests can tell which verb type shows a significant difference between children and adults.

The model terms are shown in **Table 5**. As expected, this shows a significant contrast for *have*-selecting intransitive as a main effect ($\beta = 43.23$, z = 2.87, p = 0.004). There is also a significant main effect for 'child' relative to 'adult' ($\beta = 20.31$, z = 2.17, p = 0.030) and a significant interaction for speaker (child) and verb type (*have*-selecting) ($\beta = -35.87$, z = –2.38, p = 0.017).

Fixed effect	***β***	**SE**	**z**	**p**
(Intercept)	–21.72	9.36	–2.32	0.020
verb type (*be*-intransitive)	15.20	9.35	1.62	0.104
verb type (*have*-intransitive)	**43.23**	**15.05**	**2.87**	**0.004**
speaker (child)	**20.31**	**9.36**	**2.17**	**0.030**
verb type (*be*-intransitive) : speaker (child)	–15.98	9.35	–1.71	0.088
verb type (*have*-intransitive) : speaker (child)	**–35.87**	**15.05**	**–2.38**	**0.017**

Table 5: Mixed effects logistic regression for auxiliary use in French.

In order to determine which verb type was responsible for this effect, we ran posthoc tests using the emmeans package (Lenth 2023), shown in **Table 6**. This shows that there is a significant difference in *have* productions between child and adult speakers for the reflexive ($F = 4.71$, $p = 0.030$) and *be*-selecting intransitive ($F = 54.25$, $p < 0.001$) categories, while children and adults do not differ significantly for *have*-selecting intransitives.

Turning to Italian child speech (**Figure 2**, right), there are a small number of errors, with error rates of 3% for *be*-selecting intransitive verbs (4 errors out of a total of 120 auxiliary–participle pairs), 2% for reflexives (1 error out of 56 pairs), and 1% for *have*-selecting intransitives (1 error out of 100 pairs). See the relevant child utterances in (14). (14a) to (14c) are auxiliary errors with

verb type	F	p
reflexive	**4.71**	**0.030**
***be*-intransitive**	**54.26**	**<0.001**
have-intransitive	0.79	0.375

Table 6: Posthoc tests for interaction of speaker (child/adult) by verb type.

be-selecting intransitive verbs produced by Diana (all at 2;05). (14d) is a production of *have* with a reflexive, also for Diana (2;06). (14e) uses *have* with a *be*-selecting intransitive (fall), (Raffaello, 2;01)[11] while (14f) is the only instance of *be* with a *have*-selecting intransitive (or null-object transitive) (Marco, 2;01).[12]

(14) a. ancora un po', eccolo **hai** venuto tu!
again a bit there have.2SG come you
'a bit further, there, you have come here.'

b. Ecco è [//] **ha** già **arrivato** ecco
there be.3SG … have.3SG already arrived there
'There, he has already arrived.'

c. **Ha** già **arrivato** questo
have.3SG already arrived that
'that has already arrived.'

d. Guarda, **s'ha** **messo** i ciuccio!
look REFL-have.3SG put the dummy
'look, she has put the dummy in'

e. **a** **cascato**!
have.3SG fallen
'It fell!'

f. Non **ho** **fatto**, non **sono** **fatto**.
not have.1SG done, not be.1SG done
'I didn't do it.'

It is worth noting that 4 of the 6 errors in Italian were produced by one child (Diana) and most children only produced adultlike auxiliaries. In contrast, in child French we measured at least one auxiliary error for each of the 18 French-speaking children though there were considerable

[11] This auxiliary could be considered ambiguous, as suggested by the transcription. However, it was coded as an auxiliary error here as it is followed by parental correction: *è cascato* 'it fell').

[12] This instance could be considered word play of some kind as the adultlike auxiliary is also produced beforehand.

differences in error rates by child, as well as differences between intransitives and reflexives for instance for many of the children. See supplementary files for counts by individual child.

To determine whether these error frequencies are different to those seen in French, we ran a mixed effects logistic regression model comparing auxiliary productions in child French and child Italian. This analysis takes *have* production as a binary dependent variable, with language (French/Italian) and verb type (*be*-selecting intransitive, reflexive, *have*-selecting intransitive) as fixed effects. Individual child and lexical verb were added as random intercepts.[13]

As *have*-production is the dependent variable, we expect to see a main effect of verb type (*have*-selecting intransitive). A main effect of language is also likely and would reflect different proportions of *be*-selecting verbs. More importantly, an interaction of verb type and language for any of the three verb categories would indicate a significant difference between the two languages for that verb category. Therefore, if the rate of error is higher in child French for *be*-selection alone, we would expect to see an interaction for *be*-selecting intransitive and/or reflexive, but not *have*-selecting intransitive, as French-speaking children do not produce significant errors with the latter.

The model is shown in **Table 7**. As predicted, there is a significant contrast for verb type (*have*-selecting) as a main effect ($\beta = 10.05$, $z = 7.63$, $p < 0.001$) and language (Italian) ($\beta = -2.53$, $z = -2.11$, $p = 0.035$).

Fixed effect	**β**	**SE**	**z**	**p**
(Intercept)	–3.11	0.66	–4.69	<0.001
language (Italian)	**–2.53**	**1.20**	**–2.11**	**0.035**
verb type (*be*-selecting)	–0.06	0.68	–0.09	0.925
verb type (*have*-selecting)	**10.05**	**1.32**	**7.63**	**<0.001**
language (Italian) : verb type (*be*-selecting)	1.05	1.25	0.84	0.400
language (Italian) : verb type (*have*-selecting)	1.53	1.90	0.81	0.420

Table 7: Mixed effects logistic regression for child *have* production in French and Italian.

We ran posthoc tests to determine if there are significant contrasts between languages for each of the three verb types. This is shown in **Table 8**. The analysis shows a significant interaction for reflexive ($F = 4.43$, $p = 0.035$) and *be*-selecting intransitive ($F = 4.39$, $p = 0.036$) categories, but not for *have*-selecting intransitives ($F = 0.43$, $p = 0.510$), which is consistent with a difference

[13] For the purpose of this analysis, Italian verbs were translated to a French equivalent in order to have comparable levels of the variable between languages. A model directly comparing child and adult Italian failed to converge.

between French and Italian children's use of auxiliaries with the two *be*-selecting verb categories, but not for *have*-selecting intransitives.

verb type	F	p
reflexive	**4.43**	**0.035**
***be*-selecting**	**4.39**	**0.036**
have-selecting	0.43	0.510

Table 8: Posthoc tests for language (Italian/French) by verb type.

The results show that French-speaking children produce significantly more auxiliary errors with the *be*-selecting verb categories, confirming that the much higher error frequency in French *be*-selecting verbs (10% to 20%) compared to Italian (2%–3%) is a significant difference between languages.

In the following section we examine the French auxiliary errors produced in more detail. As discussed above (section 1.3.1), a prior study reported different rates of error for reflexives by person/number: highest for 1SG, lower for 2SG, and at zero for 3SG. We therefore compare the distribution of person/number forms in both *be*-selecting verb types to better understand this effect and determine if it applies to *be*-selecting intransitives in addition to reflexives.

3.3 Auxiliary productions by person/number

Further analyses are needed to better understand auxiliary error patterns in child French. For instance, though all French children produce non-adultlike auxiliaries to some degree, there are considerable differences across individuals, and error rates are not necessarily similar between *be*-selecting intransitive and reflexive verbs. Also, it is noteworthy that there are significant errors with verbs that have shown very limited variability in the literature, such as *aller* 'go'.[14]

3.3.1 Auxiliary productions in French child speech

To gain a better understanding of the reasons children are producing such errors we analysed the specific adultlike (*be*) and non-adultlike (*have*) auxiliary forms used by each of the French-speaking children for the *be*-selecting verb categories. We focus on present tense auxiliaries (equivalent to present perfect verbs) as these make up the vast majority of perfect auxiliaries in the corpora (at 94% of *be*-selecting verbs in parental speech, and 96% in child speech) and start being produced by children before other auxiliary forms.

It is necessary to consider the verbal paradigm for present tense *be*, shown in **Table 9**. For both auxiliaries, 2SG and 3SG forms are identical single vowels, with distinct forms for 1SG

[14] See supplementary tables for auxiliary error counts by verb.

and 3PL. 1PL ‘nous’ is not typically used in colloquial and child-directed speech, and 2PL ‘vous’ is infrequent.[15]

	person/number	subject auxiliary
BE	1st-SG	je suis
	2nd-SG	tu es [e]
	3rd-SG	il est [e]
	3rd-PL	ils sont
HAVE	1st-SG	j’ai [e]
	2nd-SG	tu as [a]
	3rd-SG	il a [a]
	3rd-PL	ils ont
REFL-BE	1st-SG	je me suis
	2nd-SG	tu t’es [te]
	3rd-SG	il s’est [se]
	3rd-PL	ils se sont

Table 9: French *be* and *have* paradigms.

We first analyse the distribution of person-number forms in relation to child errors. **Figure 3** shows the total counts of each of these person/number auxiliary forms for *be*-selecting intransitive and reflexive verbs, comparing adultlike (blue) to non-adultlike (orange).[16]

It is clear from the relative proportions that there is an increased rate of error with 1SG. In the case of reflexives it is the only error observed. With *be*-selecting intransitives, while 3SG and 3PL errors are present, the relative rate of error is much higher for 1SG than 3SG. Specifically, for this verb type the rate of error for 1SG auxiliaries is 42% (42 errors out of 100 counts) while for 3SG auxiliaries it is 5.3% (35 out of 664 counts). No errors were observed for 2SG (out of 11 counts for *be*-selecting intransitives and 14 for reflexives). These results differ from those of Boyce et al. (2017), as we show that the increased rate of 1SG errors is not limited to reflexives but rather is a characteristic of both *be*-selecting verb types.

In order to determine if this asymmetry in error frequencies is significant, we ran a mixed effects regression model on the data for child productions of *be*-selecting verbs, with production of auxiliary *have* as the binary dependent variable, fixed effect for ‘person’ (1SG, 2/3SG and 3PL),

[15] There were fewer than 10 occurrences of 2PL auxiliary BE (êtes) in the parental speech as a whole in the selected French corpora.

[16] The counts by child for errors by person/number form are available in the supplementary files.

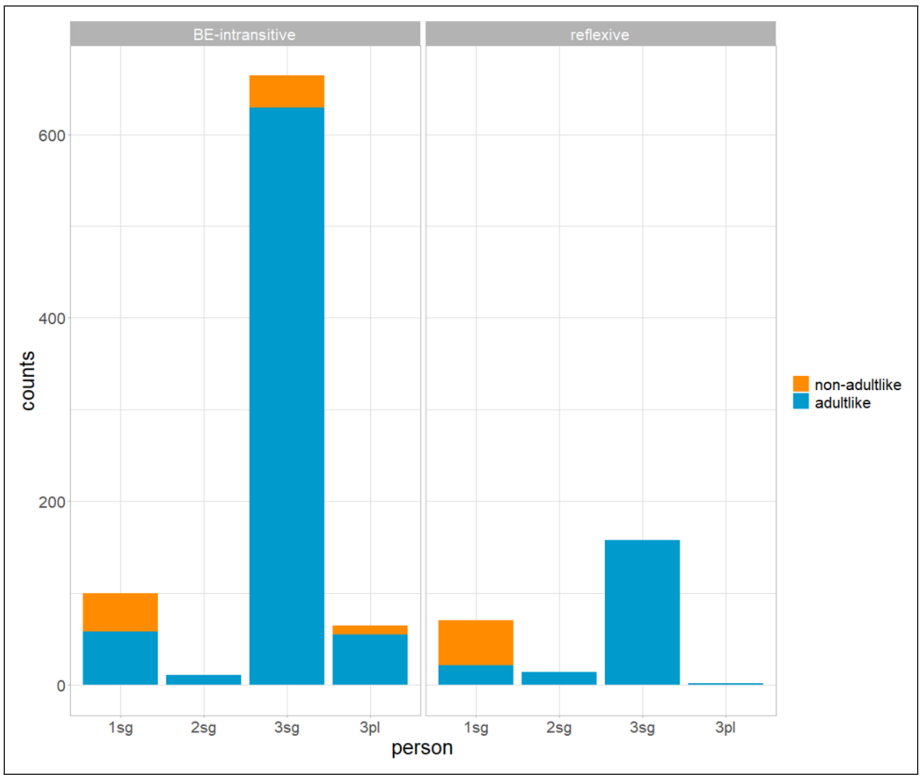

Figure 3: Person/number distribution of present tense auxiliaries in BE-selecting verbs in child French.

and both 'child' and 'verb' as random intercepts. For this analysis 2SG auxiliaries were recoded as 3SG to reflect the syncretism and homophony. A main effect of 'person' in this model would reflect a difference between any of the 1SG, 2/3SG and 3PL factors. The model terms are shown in **Table 10**. This shows an effect of the variable 'person' for 3PL ($\beta = 1.93$, z = 8.49, $p < 0.001$).

Fixed effect	***β***	**SE**	**z**	**p**
(Intercept)	–1.94	0.45	–4.33	<0.001
person (3PL)	**1.93**	**0.23**	**8.49**	**<0.001**
person (2/3SG)	0.14	0.29	0.50	0.621

Table 10: Mixed effects logistic regression for *have* production by person/number form in child French.

As we aim to determine if the asymmetry is significant between the three person/number forms, we ran pairwise tests on this model using the 'emmeans' command from the emmeans package (Lenth 2023), in order to analyse the contrasts for each pair, shown in **Table 11**.

contrast	β	SE	z	p
1SG–3PL	**1.78**	**0.46**	**3.83**	**<0.001**
1SG–2/3SG	**3.99**	**0.34**	**11.56**	**<0.001**
3PL–2/3SG	**2.21**	**0.46**	**4.83**	**<0.001**

Table 11: Pairwise comparisons of person/number levels in the child French *have*-model.

Importantly, this shows that there is a significantly higher production of *have* with 1SG than 2/3SG auxiliaries ($\beta = 3.99$, $z = 11.56$, $p < 0.001$). *Have* productions are also significantly higher with 1SG than 3PL ($\beta = 1.78$, $z = 3.83$, $p < 0.001$) and greater with 3PL compared to 2/3SG ($\beta = 2.21$, $z = 4.83$, $p < 0.001$).

In summary, this analysis shows that auxiliary errors are not evenly distributed among person/number forms. Instead, errors are more common for 1SG than other forms of *be*, with a much higher percentage of adultlike auxiliaries produced for 3SG. Furthermore, in contrast to Boyce et al. (2017) we found no errors with 2SG auxiliaries, showing that this increased error rate is specific to 1SG.

In the following sections, we consider a potential explanation for these errors due to the homophony between 1SG *have* and 3SG *be* in spoken French. Specifically, since both are realised as [e], it is possible that 1SG *be* errors are not true auxiliary selection errors (i.e. not 1SG *have*) but represent instead the extension of 3SG *be* where 1SG is required. In the following section, we analyse the frequencies of the different *be* auxiliaries in parental speech, as a high proportion of 2/3SG *be* could be a potential factor responsible for these over-productions.

3.3.2 Auxiliary use by person/number in parental speech

As there are significantly higher rates of error in child auxiliary productions for 1SG compared to other person/number forms, we analyse the distribution of these person/number forms in parental speech as a potential source of error. Importantly, as 2/3SG *be* (es/est) and 1SG *have* (ai) are both realised as [e] in spoken French, there are two potential sources for child errors with 1SG *be*: children may be selecting the wrong 1SG auxiliary (*have*) or they may be selecting the expected auxiliary (*be*), but using a 3SG 'est' instead of 1SG 'suis' (see **Table 9** above). For consistency we will refer to the former as auxiliary selection errors and the latter as 3SG extensions.

We analyse parental speech in order to determine the relative frequencies of 2/3SG [e] compared to the other forms in the *be* paradigm. If there is a significantly higher frequency of one form relative to others in the language input this could be a factor explaining why children over-produce a 3SG instead of a 1SG *be* auxiliary. As both 2SG and 3SG are realized as [e] we have combined them in this analysis.

The total counts by person/number category for reflexive and *be*-selecting intransitive verbs in French parental speech are shown in **Figure 4**.

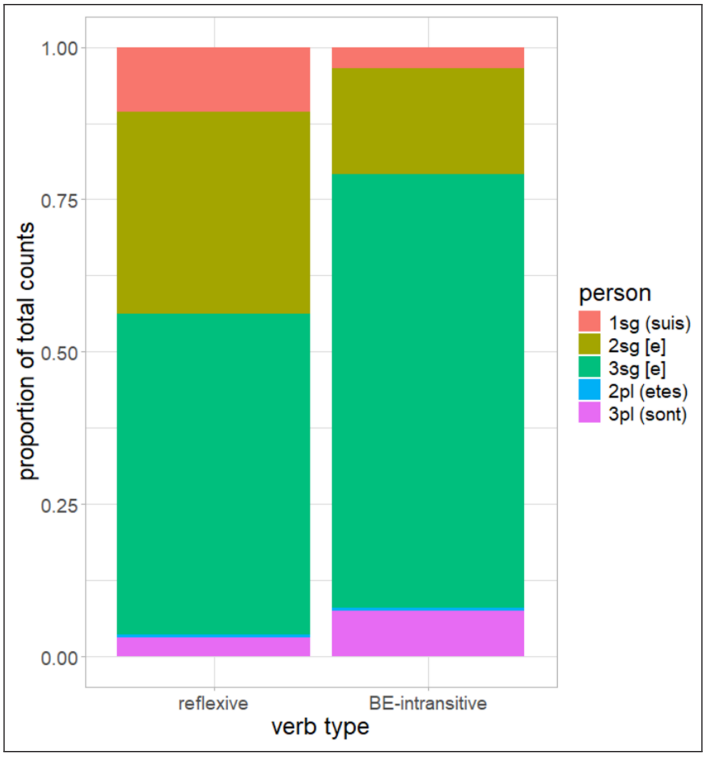

Figure 4: Person/number distribution of present tense auxiliaries in BE-selecting verbs in French parental speech (2SG and 3SG are both [e]).

In both verb types, 2SG/3SG [e] make up the vast majority of auxiliaries, at 88% of total auxiliaries for *be*-selecting intransitives (1430 out of 1617 counts) and 86% for reflexives (699 out of 816 counts), with 1SG and 3PL forms together accounting for the remaining 12% of *be*-selecting intransitive auxiliaries (57 counts for 1SG and 122 counts for 3PL out of a total of 1617) and 14% for reflexive auxiliaries (87 counts for 1SG and 26 counts for 3PL out of a total of 816).[17]

To analyse these distributions we ran a mixed effects regression model with 'counts' as a continuous dependent variable, fixed effect for 'person' (1SG, 2SG/3SG, 3PL), and both 'child' and 'verb' as random intercept terms.[18] In this case, a main effect of 'person' would be consistent with significant differences between counts for the different person/number forms.

The model terms are shown in **Table 12**. This shows a significant effect of the variable 'person' for 2SG/3SG ($\beta = 4.01$, $t = 6.34$, $p < 0.001$) but not for 3PL.

[17] 2PL forms represented less than 1% of total auxiliaries, with 4 occurrences of 2PL reflexive 'vous vous êtes' and 8 occurrences of 'vous êtes' with *be*-selecting intransitives in parental speech.

[18] 2PL auxiliaries were ignored for the purpose of this analysis.

Fixed effect	β	SE	t	p
(Intercept)	–1.60	0.73	–2.20	0.030
person (3PL)	–0.52	0.82	–0.64	0.526
person (2SG/3SG)	**4.01**	**0.63**	**6.34**	**<0.001**

Table 12: Mixed effects regression for counts by person/number form in adult French.

As for the child data, we ran pairwise tests on this model in order to analyse the contrasts for each pair, shown in **Table 13**. The contrasts show significantly lower counts of 1SG *be* compared to 2SG/3SG (β = –4.01, t = –6.33, p < 0.001) and significantly lower counts of 3PL *be* compared to 2SG/3SG (β = –4.54, t = –7.32, p < 0.001), while the contrast between 1SG and 3PL was not significant.

contrast	β	SE	t	p
1SG–3PL	0.52	0.82	0.63	0.802
1SG–2SG/3SG	**–4.01**	**0.63**	**–6.33**	**<0.001**
3PL–2SG/3SG	**–4.54**	**0.62**	**–7.32**	**<0.001**

Table 13: Pairwise comparisons of person/number levels in adult French.

In summary, there is a significant asymmetry in the distribution of person/number *be* auxiliaries in parental speech, with much higher frequencies of 2SG/3SG [e]. This confirms the skewed nature of the input, which could therefore be a factor explaining children's over-production of 3SG auxiliaries in the early stages of acquisition.

3.3.3 Auxiliary use in Italian by person/number

This section analyses the relative frequencies of person/number forms in the Italian data for child and parental speech. Mainly, if there is a similar asymmetry between the different person/number forms in the parent's speech as was found in French, this could suggest that the relative frequencies of person/number forms are not a significant factor responsible for the 1SG errors in child French, since this error type is specific to French (Italian auxiliary errors included one 2SG error with *be*, four 3SG errors with *be*, and one 1SG error with *have*, section 3.2).

The Italian paradigm for present tense *be* and *have* is shown in **Table 14**. The only syncretism is in the *be* forms for 1SG and 3PL.[19] There is no homophony between any of the *have* and *be* forms, in contrast to French (section 3.3.1).

[19] Although the 1SG and 3PL BE are both 'sono', the participle forms are distinct due to subject agreement.

	person/number	auxiliary
BE	1SG	sono
	2SG	sei
	3SG	è
	1PL	siamo
	2PL	siete
	3PL	sono
HAVE	1SG	ho
	2SG	hai
	3SG	ha
	1PL	abbiamo
	2PL	avete
	3PL	hanno
REFL-BE	1SG	mi sono
	2SG	ti sei
	3SG	si è
	1PL	ci siamo
	2PL	vi siete
	3PL	si sono

Table 14: Italian *be* and *have* paradigms.

Figure 5 shows the relative proportion of each *be* auxiliary by verb type for child and adult Italian.

As for French, we only consider present tense auxiliaries. In the parents' speech, the relative proportions are similar for both verb types (reflexive and *be*-selecting intransitives), with the majority of auxiliaries as 3SG forms, which account for 65% of *be*-selecting intransitives (555 out of a total of 849 counts) and 63% (164 out of a total of 262 counts) of reflexive verb counts respectively. In contrast, 1SG/3PL 'sono' accounts for 7% of *be*-selecting intransitives (60 out of 849) and 10% (27 out of 262) of reflexive verb counts, while 2SG 'sei' auxiliaries make up 18% of *be*-selecting intransitive pairs (150 out of 849 counts) and 27% of reflexive pairs (71 out of 262 counts). There are also 82 counts of 1PL 'siamo' with *be*-selecting intransitive verbs, representing 10% of total *be*-selecting intransitive auxiliaries.[20] On the whole these frequencies show that there

[20] There are only 2 occurrences of 2PL 'siete' with *be*-selecting intransitives and none with reflexives.

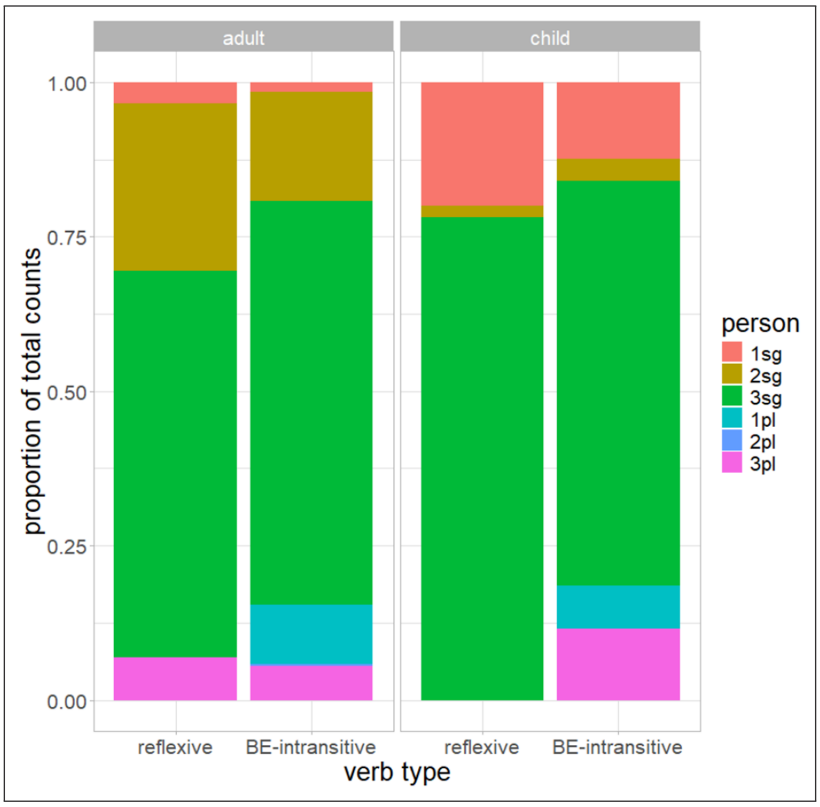

Figure 5: Person/number distribution of present tense BE auxiliaries in Italian parent and child speech (1SG and 3PL are both 'sono').

is a marked asymmetry in relative frequencies in parental speech, though possibly not to the same extent as was seen for French.

In child speech in comparison, there is a higher proportion of 1SG/3SG auxiliary 'sono' and a lower proportion of 2SG auxiliaries with both verb types. There is also a somewhat higher percentage of 3SG forms with reflexives.

To determine if the asymmetry in parental speech is significant, we conducted a mixed-effects regression model with total counts as a continuous dependent variable, person (1SG/3PL, 2SG, 3SG, 1PL) as fixed effect, and both 'child' and 'verb' as random intercepts. A main effect of 'person' would be consistent with significant differences between counts for the different person/number forms.

The model terms are shown in **Table 15**. This shows a significant effect of the variable 'person' for 2SG (β = –1.44, t = –2.49, p = 0.013).

We also conducted pairwise comparisons in order to analyse each of the person/number contrasts. The analysis of the French data showed that there was a significant contrast between the counts of 2SG/3SG and the other two auxiliary forms (1SG and 3PL) (section 3.3.2). If

Fixed effect	β	SE	t	p
(Intercept)	1.00	0.61	1.64	0.118
person (1SG/3PL)	–0.13	0.86	–0.15	0.881
person (2SG)	**–1.44**	**0.58**	**–2.49**	**0.013**
person (3SG)	–0.21	0.47	–0.41	0.684

Table 15: Mixed effects regression for counts by person/number forms in adult Italian.

the distribution of person/number auxiliaries is essentially the same in the two languages, we would expect to see similar contrasts between 3SG and the other person/number categories in Italian.[21]

The result of this analysis is shown in **Table 16**. The contrasts show significantly lower counts for 1SG/3PL *be* compared to 3SG ($\beta = -3.22$, $t = -4.31$, $p < 0.001$) and significantly lower counts of 2SG *be* compared to 3SG ($\beta = -1.99$, $t = -3.03$, $p = 0.014$). Other contrasts were not significant.

contrast	β	SE	t	p
1PL–(1SG/3PL)	1.31	1.29	1.02	0.738
1PL–2SG	0.08	1.23	0.07	0.999
1PL–3SG	–1.91	1.18	–1.62	0.369
(1SG/3PL)–2SG	–1.23	0.84	–1.46	0.461
(1SG/3PL)–3SG	**–3.22**	**0.75**	**–4.31**	**<0.001**
2SG–3SG	**–1.99**	**0.66**	**–3.03**	**0.014**

Table 16: Pairwise comparisons of person/number levels in adult Italian.

In summary, the analysis of the distribution of person/number forms in Italian parental speech shows a similar pattern to what was found for French. Although the asymmetry is more pronounced in French, with somewhat higher overall percentages of 2SG/3SG in that language compared to 3SG in Italian, in both languages there is a contrast between 3SG and the other forms of the *be* paradigm. Since Italian children do not replace 1SG *be* 'sono' with 3SG 'è', the results obtained for Italian parental speech suggest by extension that this is not likely to be the main reason French-speaking children produce non-adultlike 1SG *be* forms and therefore other underlying reasons must be responsible for these productions.

[21] We are unable to compare the two languages directly due to the differences in syncretism which is taken into account in these analyses.

If child errors with *be* auxiliaries are extensions of 3SG forms due for instance to a delayed acquisition of 1SG *be* 'suis', instead of auxiliary selection errors, we would expect this delay to apply to any production of 1SG *be*, i.e. including instances of *be* as a copula. In the following section we consider this possibility by comparing the earliest occurrences of *be* as auxiliary and copula.

3.4 Children's first productions of 1SG 'suis' as auxiliary and copula

As child error rates for French auxiliaries with *be*-selecting intransitive verbs are significantly higher for 1SG *be* '(je) suis' compared to 3SG '(il/elle) est' (section 3.3.1), it is possible that this represents a more general difficulty in the acquisition of the *be* inflectional paradigm, which would not be limited to *be* as an auxiliary. In other words, if the issue is with acquisition of the paradigm we should find a similar effect with production of 'suis' in copular constructions.

To test this hypothesis, we searched for the first occurrences of 3SG and 1SG forms for each of the children, comparing *be* and *have* for uses as both auxiliaries and copula/main verb, as a measure of the relative delay in production of 1SG compared to 3SG forms.[22] If the difficulty in acquisition is limited to *be* as an auxiliary (and the errors with 1SG *be* are due to auxiliary selection), we would expect to see a delay in the first occurrences of '(je) suis' as an auxiliary compared to '(je) suis' as copula, relative to first occurrences of 3SG '(il/elle) est'. If on the other hand the difficulty is due to the acquisition of the *be* paradigm (and the errors with 1SG *be* are due to extensions of 3SG *be*) we would expect to find a similar delay in production of 1SG *be* independently of type (auxiliary or copula), relative to first productions of 3SG *be*. We also measured first occurrences of 1SG and 3SG *have* as auxiliary and main verb to determine if this possible 1SG delay is limited to *be*.

The results of this analysis are shown in **Figure 6** for *be* (left) and *have* (right), comparing 1SG and 3SG as auxiliary or copula/main verb, plotted by age in months. Comparing the median age of first occurrence for *be*, there is a general delay between 1SG and 3SG forms which does not appear to depend on whether this is an auxiliary or copula i.e. there is a similar delay for 1SG relative to 3SG in both types. In comparison, for *have* there appears to be no relative delay (1SG vs. 3SG) for the auxiliary and a slight delay for the main verb.

To test the significance of these results we conducted a mixed effects regression model with age in months as the dependent (continuous) variable and the three independent variables as fixed effects: verb (be/have), type (auxiliary/main verb), person (1SG/3SG). 'Child' and 'verb' were fitted as random intercepts. For either verb (*be* and *have*), a main effect of 'person' would indicate a significant delay in the occurrence of 1SG compared to 3SG, independently of 'type',

[22] Although this analysis would also be possible for Italian, it requires a sufficient number of consistent recording dates. We believe that the data available for Italian is too limited to make this particular analysis valid for this language.

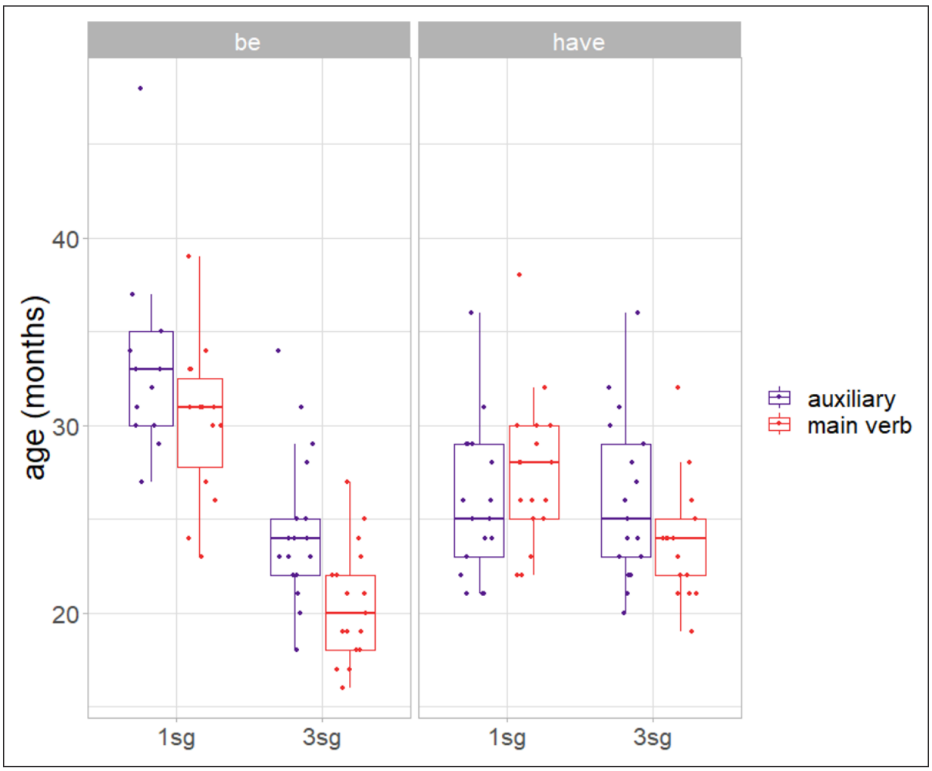

Figure 6: First occurrence of 1SG and 3SG BE and HAVE as auxiliary and main verb in child speech.

while a main effect of 'type' (auxiliary/main verb) would reflect a delay in the first occurrence of either (independently of 'person'). An interaction between 'person' and 'type' would indicate a greater delay in the production of the 1SG in either type (auxiliary or main verb). Crucially, if the difficulty in the acquisition of 1SG *be* is due to auxiliary selection, we would expect an increased delay with 1SG auxiliaries compared to 1SG copulas, i.e. an interaction between 'type' and 'person' for *be*. If on the other hand it is related to an inherent difficulty in acquiring the full *be* paradigm, we would expect a similar delay for 1SG relative to 3SG independently of 'type' i.e. no interaction between these variables. We make no specific predictions for *have*, though the absence of 1SG errors with auxiliary *have* would be consistent with a different pattern for this verb reflecting the absence of a similar delay.

The model fixed effects are shown in **Table 17**. There is a significant effect of person (3SG) (β = –8.90, t = –9.52, p < 0.001) reflecting on average earlier production of 3SG forms (by approximately 9 months), a significant main effect of verb *have* (β = –7.48, t = –8.01, p < 0.001) reflecting an average earlier production of this verb (by approximately 7.5 months), and a significant effect of type (main verb) (β = –3.09, t = –3.16, p = 0.002) consistent with an earlier production of main verb forms (by approximately 3 months). There is also a significant

Fixed effect	β	SE	z	p
(Intercept)	33.37	1.03	32.38	<0.001
person (3SG)	**–8.90**	**0.93**	**–9.52**	**<0.001**
verb (have)	**–7.48**	**0.93**	**–8.01**	**<0.001**
type (main verb)	**–3.09**	**0.98**	**–3.16**	**0.002**
person (3SG) : verb (have)	**9.07**	**1.27**	**7.14**	**<0.001**
person (3SG) : type (main verb)	–0.91	1.30	–0.70	0.488
verb (have) : type (main verb)	**4.74**	**1.30**	**3.63**	**<0.001**
person (3SG) : verb (have) : type (main verb)	–2.97	1.79	–1.67	0.099

Table 17: Mixed effects logistic regression for age of first occurrence of 1SG and 3SG *be* and *have* auxiliary and main verb forms in child French.

interaction of person (3SG) with verb (*have*) ($\beta = 9.07$, $t = 7.14$, $p < 0.001$), and an interaction between verb (*have*) and type (main verb) ($\beta = 4.74$, $t = 3.63$, $p < 0.001$).

We ran posthoc tests by verb (*be/have*) on this model, shown in **Table 18** for *be* and **Table 19** for *have*. For *be* there is a main effect of 'person' (1SG/3SG) ($F = 204.57$, $p < 0.001$) and 'type' (auxiliary/main verb) ($F = 29.53$, $p < 0.001$) but no interaction between these variables. These results show that there is a general delay for 1SG (relative to 3SG) *be* and a general delay of *be* as an auxiliary (independently of person/number), but importantly, no increased delay for 1SG *be* as an auxiliary. This analysis is therefore consistent with a general delay for 1SG *be*, supporting the hypothesis that 1SG errors represent a difficulty in the acquisition of the *be* paradigm, rather than a difficulty specific to the acquisition of *be*-auxiliary selection.

variable	F	p
person (1SG/3SG)	**204.57**	**<0.001**
type (auxiliary/main verb)	**29.53**	**<0.001**
person : type	0.48	0.488

Table 18: Posthoc test for first occurrence of *be* by person and type.

variable	F	p
person (1SG/3SG)	**8.39**	**0.005**
type (auxiliary/main verb)	0.23	0.630
person : type	**10.15**	**0.002**

Table 19: Posthoc test for first occurrence of *have* by person and type.

For *have* first occurrences there is a main effect of person ($F = 8.39$, $p = 0.005$), reflecting a somewhat later occurrence of 1SG overall, but no main effect of type (auxiliary/main verb). The interaction between these two variables is significant ($F = 10.15$, $p = 0.002$), which can be interpreted as an increase in the delay for the first occurrence of 1SG with one of the two types (main verb as shown in **Figure 6**).

To summarize our results for French, there are significant errors in child French *be*-selecting intransitives and reflexive verbs. These are not due to input variability given that there are no non-standard uses in French parental speech. Furthermore, these errors occur preferentially with 1SG *be*. Although there are two possible sources of error for 1SG *be*, our results show that the errors are likely to represent the delayed acquisition of the 1SG form of the *be* paradigm and are not limited to auxiliary uses. The asymmetric nature of the input by person/number, with a much greater frequency of 2SG/3SG *be* in parental speech may be a factor though other factors are likely to be responsible as similar errors do not occur in child Italian.

4 Discussion

We first aimed to better characterize child auxiliary selection errors. This was important since determining whether these errors are independent of or restricted to *be*-selection has implications for how auxiliary selection is initially acquired and represented. More specifically, if errors are restricted to *be*-selecting categories this would be consistent with children initially acquiring and generalising *have* as a perfective auxiliary and subsequently acquiring *be*-selection for the subset of verbs that require it, whereas if errors are evenly distributed this would suggest that both auxiliaries are acquired simultaneously. Our analysis of child French (section 3.2) indeed shows that these errors are not evenly distributed across the three verb types studied, but are instead specific to *be*-selecting verbs: children produce non-adultlike auxiliaries at a significant rate with *be*-selecting intransitives and reflexives but not with *have*-selecting intransitives.[23] In this regard our results differ from those of Boyce et al. (2017), who did not find significant differences in error rates between *be*- and *have*-selecting intransitives. This difference in results is somewhat hard to explain. One possible explanation is that many of the most common *have*-selecting past participles can take resultative meanings, such as *finir* 'finish' and *gagner* 'win', and that instances of *be*-intransitives were excluded in our analysis if it could be determined that they were used in

[23] The two recorded instances of auxiliary *be* with *have*-selecting intransitive verbs were the following:

(i) a. Après il est couru jusque …('after, he is run until …', Anaé, 3;10)
b. On est disparu avec Luc et Luc est dans son lit. ('we are disappeared with Luc and Luc is in his bed') (Léa, 2;09)

this way.[24] Additionally, some instances of apparent 3SG *be* with *have*-selecting intransitives were excluded based on the audio recording if the auxiliary was not sufficiently audible. Our results for French are however in line with previous studies that found that errors were restricted to *be*-selection (Snyder et al. 1995).

The results obtained for Italian (section 3.2) show that although there are some auxiliary errors at a low frequency in this language, these are not specific to *be*-selection: in Italian, children do not make errors at a significantly higher frequency with *be* auxiliaries. This is supported by the regression model which shows significant differences in *have* productions between languages for *be*-selecting intransitive and reflexive verbs. It is worth noting that this model required translating the Italian verbs into French equivalents in order to have comparable levels of the variable. This approach is not entirely objective as in some cases the closest equivalent verb in meaning is not of the same type, such as a verb which is intransitive in one language but reflexive in the other (e.g. *dimenticarsi/oublier* 'forget'), or for instance the fact that a verb like French *tomber* 'fall' has two synonyms in Italian (*cascare* and *cadere*).

Furthermore, we have shown through the analysis of parental speech that auxiliary selection errors in child French are not due to variability in the language input (section 3.1). This is true for all three verb categories (reflexive, *be*-selecting and *have*-selecting intransitives), for which the auxiliaries produced in parental speech match those expected for standard French. It is possible that this lack of variability represents the influence of socio-economic factors, the location of the participants, the limited influence of contact with English, or a combination of these factors.

As several studies have shown the presence of variability in auxiliary selection in at least some varieties of spoken French (Sankoff & Thibault 1977; Russo & Roberts 1999; King & Nadasdi 2005; Rea 2020), without studying parental speech we would have been unable to know for certain whether the observed errors originate from the child or are merely the result of input variability. Our results confirm that the errors are indeed child innovations reflecting in some way the nature of the acquisition process. Our next aim was to better understand the nature of these errors.

Our analysis of present tense *be* auxiliaries (section 3.3.1) shows that errors are not evenly distributed across all person/number forms, but rather are much more frequent for 1SG than 3SG auxiliaries. In Boyce et al.'s study (2017) this effect was reported but was considered to be specific to reflexive verbs and therefore a distinguishing factor between the two verb types.

[24] An example of this is the following which initially appears to be intransitive as the child uses the verb 'push', but when studied in context is clearly adjectival as she is talking about a teddy bear's head that was squashed when someone sat on it:

(i) Il est appuyé un peu dessus. ('he is pushed (squashed) a bit on top', Madeleine: 2;06)

They also postulated that the syncretism between reflexive and object clitics (for 1SG and 2SG but not 3SG) could be a factor leading to non-adultlike auxiliary productions, as children could be incorrectly interpreting these constructions as transitive (section 1.3). However, our results show that this effect is not limited to reflexives since there is a significantly higher frequency of 1SG auxiliary errors with both reflexives and *be*-selecting intransitives. Another difference between our results and those of Boyce et al. (2017) concerns potential errors with 2SG reflexives (i.e. *tu t'as* + participle). They report a small number of such errors, while we did not find any (in a total of 14 occurrences of adultlike *tu t'es*): only 1SG errors were observed with reflexive verbs. We conclude from this that the difficulty in production of 1SG 'je suis' applies to *be* auxiliaries in general and is independent of reflexive clitic constructions.

One possible explanation for the high rate of error with 1SG *be* auxiliaries is that these are not in fact auxiliary selection errors, but rather productions of 3SG *be* 'est' in 1SG contexts. These two sources of error are indistinguishable since both 3SG *be* (est) and 1SG *have* (ai) are homophonous (realised as [e] in spoken French). In support of this, we confirmed that the language input shows a much higher frequency of 3SG *be* auxiliaries compared to other person/number forms (section 3.3.2), which is enhanced in spoken French in part due to the syncretism between 2SG and 3SG forms together with the replacement of 1PL 'nous sommes' with the 3SG 'on est'.[25] The regression model shows that there are significantly higher numbers of 2SG/3SG auxiliaries compared to either 1SG or 3PL in parental speech. One possible developmental trajectory which is consistent with the asymmetry observed in the language input is that children initially learn the 3SG form of the auxiliary and treat it as a default, before the other forms in the paradigm are acquired. This would suggest that 3SG is the base or unmarked form in the paradigm, rather than the infinitive as sometimes assumed (Bybee 1991).

Further support for this account can be found in other contexts where extensions of 3SG verbs are known to occur, both in French and in other languages (Bybee 1991). An example of this is in 1st person clefts in some varieties of French (as in English, though not in standard French) (15) where the a 3SG auxiliary *a* replaces the 1SG *ai*.

(15) C'est moi qui **a** gagné.
DEM.3SG-be.3SG me that have.3SG won.
'it's me who has won.'

Examples from the data considered in this study include 3SG extensions with transitive verbs, such as (16) (Madeleine, 2;01) though we did not specifically measure the rate of occurrence of such errors.

[25] Excluding a small number of occurrences in songs and books, there were only three occurrences of 1PL BE ('nous sommes') in parental speech in the corpora.

(16) Moi l'**a** mangé.
me ACC.3SG-have.3SG eaten.
'I've eaten it.'

Similarly, though this is not specific to the production of auxiliaries, in languages such as Modern Greek which lack an infinitive, 3SG forms are considered unmarked and have been shown to be extended to other person/number contexts during early child acquisition, while extensions of other person/number forms are rare (Varlokosta et al. 1998).

Observations such as these support the analysis of 1SG *be* errors as extensions of 3SG forms in first person contexts, but it is also important to know whether this effect is specific to auxiliaries or applies to *be* generally. Indeed if it is a general characteristic of the acquisition of the *be* paradigm there would be no reason to expect differences between uses of *be*, i.e. it should be independent of whether *be* is produced as a perfective auxiliary or as a copula. Our results on the age of first occurrence of 1SG *be* ('suis') (section 3.4) support this conclusion as the relative delay in first occurrences of 1SG compared to 3SG *be* is independent of specific use (i.e. whether *be* is an auxiliary or copula). Interestingly, this delay was specific to the production of *be* 'je suis' and was not seen for 1SG *have* 'j'ai', which suggests that it is not just the result of relative input frequencies, as we should expect similar input distributions for auxiliary *have,* due to this verb's present tense inflectional paradigm having the same syncretism as *be* (**Table 9**, section 3.3.1).

In contrast, the small number of child errors in Italian are not specific to 1SG *be*: indeed most of the errors were with 3SG auxiliaries. Italian parental speech was also essentially similar to French in following standard rules of auxiliary selection, with no observed variability in auxiliary use (section 3.1). Italian parental speech also displayed a similar marked asymmetry for 3SG *be* compared to other forms. Taken together, the analysis of the Italian data suggests that input frequencies alone cannot explain the patterns of child productions, such as the 1SG errors in French. Though input frequencies may be a factor, other morpho-phonological characteristics must play a role. One limitation of this comparison between the two languages is that there is relatively less data available for Italian-speaking children (**Table 2** and **Table 4**).

One possible cause for the relative delay in production of 1SG *be* in French is that this is due to the combination of lower input frequencies and the acquisition of a suppletive form. For instance, studying the acquisition of 'go' in English, Theakston et al. (2002) found that input frequencies largely determined the order of acquisition of the different forms of 'go' in the paradigm, with the suppletive form 'went' acquired last for almost all structures with 'go' that were measured. However, another study on the acquisition of English subject-auxiliary structures found only a partial relation between input frequencies for subject-auxiliary pairs and rates of child auxiliary omissions (Theakston et al. 2005). This study found higher rates of auxiliary omissions in child speech with 1SG and 2SG auxiliaries (both *be* and *have*), compared to 3SG auxiliaries. For instance,

on average children produced 'I'm' and 'you're' with an auxiliary in less than 40% of utterances, while their production of auxiliaries with 3SG (he's, she's, it's) was much higher at 77 to 85% of utterances. However, the role of input is difficult to compare with our study as the authors only considered the frequency of subject-auxiliary pairs: since 1SG and 2SG auxiliaries are always used with the same subject this could have underestimated the frequency of 3SG auxiliaries. Although there are limitations to measuring first occurrences of specific forms in data from corpora, as recording frequencies differ to some extent between children, for *be* the effect appears to be quite marked. It is also worth noting that the analysis does not take into account the fact that for some of the children there were no recorded occurrences of 1SG *be*: we did not find any productions of 'suis' as an auxiliary for 4 of the 17 children, and as copula for 3 of these children.

The child French data also included occurrences of (true) auxiliary selection errors with *be*-selecting intransitives. This includes 3SG and 3PL *be* errors as well as a small number of past tense auxiliaries e.g. 'avait' which are unambiguously *have* auxiliaries. It is interesting to note that these errors are only observed with *be*-selecting intransitives and do not occur with reflexives.[26] This also suggests that different pathways of acquisition are responsible for each of the two verb categories, which goes against an analysis of reflexives as unaccusatives as in Snyder et al. (1995). Specifically, the absence of 3SG errors with reflexives suggests that children acquire auxiliary selection with these verbs early, producing auxiliary *be* reliably once they start producing reflexive clitics with perfective structures (and since production of the reflexive clitic was a criteria for the searches).

On the other hand, it is likely that for French intransitive verbs children need to acquire *be*-selection item-by-item and therefore generalise auxiliary *have* with intransitive verbs for which this has not been acquired. Given sufficient input, children eventually acquire auxiliary selection rules for all *be*-selecting intransitives. It is possible however, that with input variability, i.e. less than 100% *be* for any particular verb, children do not acquire this pattern reliably, which may explain why variability is associated with some intransitives (e.g. *passer* 'pass', *descendre* 'go down') more than others (such as *aller* 'go'). This could be a factor for instance in bilingual contexts where there is proportionately less French input. Another potential factor is whether a verb can be used transitively with *have* (**Table 3**) as we could expect higher rates of auxiliary *have* with these intransitives. For *descendre*, *monter* and *passer* this is consistent since they have the highest rate of 3SG errors[27] while transitive uses account for 15–18% of occurrences. However, we observed no errors with *sortir* despite transitive uses accounting for 50% of occurrences. It is hard to tell whether the pattern of acquisition of *be*-selection observed in French also applies to

[26] We did however find one occurrence of past tense 'avais' with a reflexive:

(i) Je m'avais [étais] trompé. ('I had made a mistake'), Max (2;08).

[27] See supplementary files for rate of error by verb.

Italian due to the more limited data and the small number of child errors. For instance, further research would be needed to determine if the difference in the number of *be*-selecting intransitives in French and Italian impacts the acquisition of auxiliary selection with this verb category.

5 Conclusion

This study has shown that child errors in French *be* auxiliaries occur in the presence of a uniform linguistic input, with high-frequency 1SG errors explained as non-adultlike extensions of 3SG *be* forms in 1SG contexts. This represents a general delay in children's adultlike production of 1SG form of *be* and is therefore not due to a non-adultlike representation of the rules of auxiliary selection. While input frequencies may play a role e.g. in 3SG forms being initially acquired as an unmarked form in the paradigm, the observation that this does not occur with *have* or with *be* auxiliaries in Italian shows that other morpho-phonological factors must be involved. Future research avenues could use experimental approaches to further study the acquisition of *be* auxiliaries or study acquisition in contexts where variability in auxiliary selection has been shown to occur.

Abbreviations

ACC = accusative, DEM = demonstrative, NEG = negative, PL = plural, PST = past, REFL = reflexive SG = singular

Supplementary files

All data associated with this study is available at https://doi.org/10.17605/OSF.IO/XWYGZ. This includes supplementary results files, CLAN commands and search procedures, data sets and R code used in the analyses.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments and suggestions, which allowed us to substantially improve the paper.

Competing interests

The authors have no competing interests to declare.

References

Amato, Irene. 2023. *Auxiliary selection in Italo-Romance*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.281

Antelmi, Donella. 1998. *La Prima grammatica dell'italiano: Indagine longitudinale sull'acquisizione della morfosintassi italiana*. Bologna: Il Mulino.

Aranovich, Raúl. 2007. Split auxiliary selection from a crosslinguistic perspective. In Aranovich, Raúl (ed.), *Split Auxiliary Systems: A cross-linguistic perspective* (Typological Studies in Language), 1–23. Amsterdam/Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/tsl.69.02ara

Balcom, Patricia. 2008. On the learning of auxiliary use in the referential variety by speakers of New Brunswick Acadian French. *Canadian Journal of Linguistics* 53(1). 7–34. DOI: https://doi.org/10.1017/s0008413100000876

Bard, Ellen Gurman & Robertson, Dan & Sorace, Antonella. 1996. Magnitude estimation of linguistic acceptability. *Language* 72(1). 32–68. DOI: https://doi.org/10.2307/416793

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bentley, Delia & Eythórsson, Thórhallur. 2004. Auxiliary selection and the semantics of unaccusativity. *Lingua* 114(4). 447–471. DOI: https://doi.org/10.1016/S0024-3841(03)00068-8

Bjorkman, Bronwyn. 2011. *BE-ing default: The morphosyntax of auxiliaries*. Cambridge, MA: MIT Thesis. https://dspace.mit.edu/handle/1721.1/68911.

Borer, Hagit. 1994. The projection of arguments. *University of Massachusetts Occasional Papers in Linguistics* 20(1). 19–48. https://scholarworks.umass.edu/umop/vol20/iss1/3.

Boyce, Veronica & Aravind, Athulya & Hackl, Martin. 2017. Lexical and syntactic effects on auxiliary selection: Evidence from child French. In LaMendola, Maria & Scott, Jennifer (eds.), *Proceedings of the 41st annual Boston University Conference on Language Development*, 101–113. Somerville, MA: Cascadilla Press.

Burzio, Luigi. 1986. *Italian syntax: A goverment-binding approach*. Dordrecht: Reidel. DOI: https://doi.org/10.1007/978-94-009-4522-7

Bybee, Joan. 1991. Natural morphology: The organization of paradigms and language acquisition. In Huebner, Thom & Ferguson, Charles A. (eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories: Language Acquisition and Language Disorders*, vol. 2, 67–92. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lald.2.08byb

Canale, Michael & Mougeon, Raymond & Beniak, Edouard. 1978. Acquisition of some Grammatical Elements in English and French by Monolingual and Bilingual Canadian Students. *The Canadian Modern Language Review* 34(3). 505–524. DOI: https://doi.org/10.3138/cmlr.34.3.505

Champaud, Christian. 2004. CHILDES French Champaud corpus. DOI: https://doi.org/10.21415/T5M88F

Cipriani, Paola. 2004. CHILDES Italian Calambrone corpus. DOI: https://doi.org/10.21415/T56C8R

Demuth, Katherine & Tremblay, Annie. 2008. Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language* 35(1). 99–127. DOI: https://doi.org/10.1017/S0305000907008276

Embick, David. 2004. Unaccusative syntax and verbal alternations. In Alexiadou, Artemis & Anagnostopoulou, Elena & Everaert, Martin (eds.), *The unaccusativity puzzle: Explorations of the syntax-lexicon interface*, 137–158. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199257652.003.0006

Folli, Raffaella. 2001. *Constructing telicity in English and Italian*. Oxford: University of Oxford dissertation.

Gillmann, Melitta. 2015. Auxiliary selection in closely related languages: the case of German and Dutch. In Kailuweit, Rolf & Rosemeyer, Malte (eds.), *Auxiliary selection revisited*, 333–358. Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110348866-012

Grimshaw, Jane Barbara. 1992. *Argument structure* (Linguistic Inquiry Monographs 18). Cambridge, MA: MIT Press.

Hamann, Cornelia & Ohayon, Stéphanie & Dubé, Sébastien & Frauenfelder, Ulrich H. & Rizzi, Luigi & Starke, Michal & Zesiger, Pascal. 2003. Aspects of grammatical development in young French children with SLI. *Developmental Science* 6(2). 151–158. DOI: https://doi.org/10.1111/1467-7687.00265

Heidinger, Steffen. 2015. The auxiliary selection of French monter 'move upward' from the 16th to the 20th century. In Kailuweit, Rolf & Rosemeyer, Malte (eds.), *Auxiliary selection revisited*, 277–300. Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110348866-010

Hoekstra, Teun. 1999. Auxiliary Selection in Dutch. *Natural Language & Linguistic Theory* 17(1). 67–84. DOI: https://doi.org/10.1023/A:1006186109813

Kailuweit, Rolf. 2015. BE or HAVE in Contemporary Standard French - residua of semantic motivation. In Kailuweit, Rolf & Rosemeyer, Malte (eds.), *Auxiliary selection revisited*, 249–276. Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110348866-009

King, Ruth & Nadasdi, Terry. 2005. Deux auxiliaires qui voulaient mourir en français acadien. In Brasseur, Patrice & Falkert, Anika (eds.), *Français d'Amérique : approches morphosyntaxiques*, 103–111. Paris: l'Harmattan.

Legendre, Géraldine. 2007. On the typology of auxiliary selection. *Lingua* 117(9). 1522–1540. DOI: https://doi.org/10.1016/j.lingua.2006.06.005

Legendre, Géraldine & Sorace, Antonella. 2010. Auxiliaries and intransitivity in French and in Romance. In Godard, Danièle (ed.), *Fundamental issues in the Romance languages*, 171–220. Stanford, CA: CSLI Publications.

Lenth, Russell V. 2023. emmeans: Estimated marginal means, aka least-squares means. https://CRAN.R-project.org/package=emmeans.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analysing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Manente, Mara. 2008. *L'aspect, les auxiliaires être et avoir et l'hypothèse inaccusative dans une perspective comparative français/italien*. Venice and Paris: Università Ca' Foscari and Université de Paris VIII dissertation.

Manzini, Maria Rita. 1986. On Italian SI. In Borer, Hagit (ed.), *The syntax of pronominal clitics*, 241–262. New York: Academic Press. DOI: https://doi.org/10.1163/9789004373150_011

McFadden, Thomas. 2007. Auxiliary Selection. *Language and Linguistics Compass* 1(6). 674–708. DOI: https://doi.org/10.1111/j.1749-818x.2007.00034.x

McGinnis, Martha. 2022. Reflexive clitics are verbal, not pronominal. *Canadian Journal of Linguistics* 67(3). 328–352. DOI: https://doi.org/10.1017/cnj.2022.28

Morgenstern, Aliyah & Parisse, Christophe. 2007. Codage et interprétation du langage spontané d'enfants de 1 à 3 ans. *Corpus* 6. 55–78. DOI: https://doi.org/10.4000/corpus.922

Mougeon, Raymond & Beniak, Édouard. 1986. Le français en situation de contact et la variation linguistique : le français parlé en Ontario (Canada). In *Actes du XVII Congrès International de Linguistique et Philologie Romanes, vol. 6, Variation linguistique dans l'espace : dialectologie et onomastique*, 291–314. Aix-en-Provence: Publication de l'Université de Provence.

Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the fourth annual meeting of the Berkeley Linguistics Society*, vol. 4. 157–89. Berkeley: University of California. DOI: https://doi.org/10.3765/bls.v4i0.2198

Plunkett, Bernadette. 2002. Null subjects in child French interrogatives: A view from the York corpus. In *Romance corpus linguistics: Corpora and spoken language*, 441–452. Tübingen: Narr.

R Core Team. 2023. *R: A Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Randall, Janet H. 2007. Parameterized auxiliary selection: a fine-grained interaction of features and linking rules. In Aranovich, Raúl (ed.), *Split auxiliary systems: A cross-linguistic perspective* (Typological Studies in Language), 207–235. Amsterdam/Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/tsl.69.10ran

Rastelli, Stefano. 2023. Italian verbs with two auxiliaries: A forced-choice experiment. *Probus* 35(1). 61–98. DOI: https://doi.org/10.1515/probus-2022-0013

Rea, Béatrice. 2020. *Je m'ai fait mal quand j'ai tombé: A real- and apparent-time study of auxiliary alternation in intransitive and pronominal verbs in spoken Montréal French (1971–2016)*. Oxford: University of Oxford dissertation. http://www.admin.ox.ac.uk/statutes/352-051a.shtml.

Reinhart, Tanya & Siloni, Tal. 2005. The lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry* 36(3). 389–436. DOI: https://doi.org/10.1162/0024389054396881

Rosemeyer, Malte. 2014. *Auxiliary selection in Spanish: Gradience, gradualness, and conservation* (Studies in Language Companion Series). Amsterdam/Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/slcs.155

Russo, Marijke & Roberts, Julie. 1999. Linguistic change in endangered dialects: The case of alternation between avoir and être in Vermont French. *Language Variation and Change* 11(1). 67–85. DOI: https://doi.org/10.1017/S0954394599111049

Sankoff, Gillian & Thibault, Pierrette. 1977. L'alternance entre les auxiliaires avoir et être en français parlé à Montréal. *Langue Française* 34. 81–108. DOI: https://doi.org/10.3406/lfr.1977.4818

Schwarze, Christoph. 1996. The syntax of Romance auxiliaries. In Butt, Miriam (ed.), *The proceedings of the LFG '96 conference*, 1–20. Grenoble: Rank Xerox Research Centre.

Snyder, William & Hyams, Nina & Crisma, Paola. 1995. Romance auxiliary selection with reflexive clitics: Evidence for early knowledge of unaccusativity. In Clark, Eve V. (ed.), *Proceedings of the 26th annual Child Language Research Forum*, 127–136. Stanford, CA: CSLI Publications.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76(4). 859–890. DOI: https://doi.org/10.2307/417202

Suppes, Patrick & Smith, Robert & Leveillé, Madeleine. 1973. The French syntax of a child's noun phrases. *Archives de Psychologie* 42. 207–269.

Theakston, Anna L. & Lieven, Elena V. M. & Pine, Julian M. & Rowland, Caroline F. 2002. Going, going, gone: The acquisition of the verb 'go'. *Journal of Child Language* 29(4). 783–811. DOI: https://doi.org/10.1017/S030500090200538X

Theakston, Anna L. & Lieven, Elena V. M. & Pine, Julian M. & Rowland, Caroline F. 2005. The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics* 16(1). 247–277. DOI: https://doi.org/10.1515/cogl.2005.16.1.247

Tonelli, Livia & De Marco, Anna & Vollmann, Ralf & Dressler, Wolfgang U. 1998. Le prime fasi dell'acquisizione della morfologia. *Parallela* 6. 281–301.

van Valin, Robert D. 1990. Semantic parameters of split intransitivity. *Language* 66(2). 221–260. DOI: https://doi.org/10.2307/414886

Varlokosta, Spyridoula & Vainikka, Anne & Rohrbacher, Bernhard. 1998. Functional projections, markedness, and “root infinitives” in early child Greek. *The Linguistic Review* 15(2–3). 187–208. DOI: https://doi.org/10.1515/tlir.1998.15.2-3.187

Vernice, Mirta & Sorace, Antonella. 2018. Animacy effects on the processing of intransitive verbs: an eye-tracking study. *Language, Cognition and Neuroscience* 33(7). 850–866. DOI: https://doi.org/10.1080/23273798.2017.1421316

Willis, Lauren A. 2000. *Être ou ne plus être: auxiliary alternation in Ottawa-Hull French*. Ottawa: University of Ottawa MA thesis.

Yamaguchi, Naomi. 2012. *Parcours d’acquisition des sons du langage chez deux enfants francophones.* Paris: Paris 3 thesis. http://www.theses.fr/2012PA030048.

Zaenen, Annie. 1993. Unaccusativity in Dutch: Integrating syntax and lexical semantics. In Pustejovsky, James (ed.), *Semantics and the lexicon*, 129–161. Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-1972-6_9

Zribi-Hertz, Anne. 1987. La réflexivité ergative en français moderne. *Le français moderne* 55(1–2). 23–54.