

ARTICLE

Zooming in on the semantics of French ingressives: a collocation analysis

Filip Verroens 

Universiteit Gent (Belgium)
Email: filip.verroens@ugent.be

(Received 26 May 2023; revised 29 December 2023; accepted 08 January 2024)

Abstract

This article examines the semantic value of the infinitive in the ingressive constructions *se mettre à* (SMA) and *commencer à* (COMA) using a distinctive collexeme analysis. We find that the collexemes significant for the construction SMA are fairly homogeneous across the different corpora and can be grouped into the general category of expressive collexemes. The collexemes significant for COMA are more heterogeneous and belong to the category of cognitive collexemes and to semantic fields of sensory and creative acts. The results are compatible with the hypothesis put forward by Verroens and De Cuypere (2023) stating that the overall meaning of the SMA construction is intrinsically punctual. The punctual value of SMA is not only compatible with expressive collexemes, but, moreover, emphasizes their unforeseen and unintentional meaning. Conversely, the incremental value of COMA is consistent with the gradual onset of cognitive and sensory collexemes.

Résumé

Cet article examine la valeur sémantique de l'infinitif dans les constructions inchoatives *se mettre à* (SMA) et *commencer à* (COMA) en utilisant une analyse collocationnelle distinctive. Nous constatons que les collexèmes significatifs pour la construction SMA sont assez homogènes à travers les différents corpus et peuvent être regroupés dans la catégorie générale des collexèmes expressifs. Les collexèmes significatifs pour COMA sont plus hétérogènes et appartiennent à la catégorie des collexèmes cognitifs et aux champs sémantiques des actes sensoriels et créatifs. Les résultats sont compatibles avec l'hypothèse avancée par Verroens et De Cuypere (2023) disant que le sens global de la construction SMA est intrinsèquement ponctuel. La valeur ponctuelle de SMA est non seulement compatible avec les collexèmes expressifs, mais, de plus, souligne leur sens imprévu et involontaire. À l'inverse, la valeur incrémentale de COMA est cohérente avec le commencement graduel des collexèmes cognitifs et sensoriels.

1. INTRODUCTION¹

Both ingressive² constructions *commencer à* (COMA) ‘to begin’ + Vinf. and *se mettre à* (SMA) ‘to start’ + Vinf.³ are commonly considered to be synonyms in reference grammars (e.g. Wilmet, 1998:321) and are used interchangeably at first sight:

- (1) Bientôt, l’ETA **commence** à regarder du côté du tiers-monde, se **met** à parler de “guerre révolutionnaire”. (LM111)
 ‘Soon, the ETA begins to look towards the Third World, starts to speak of “revolutionary war”.’

Nevertheless, several researchers have noted syntactic and semantic differences between both constructions (see section 2). The analysis of the collexemes, however, still represents a research gap. The aim of the current article is to define the preferences of the ingressive verb in relation to the infinitive verb by carrying out a collostructional analysis. With this statistical technique, it will be possible to specify which categories of verbs are particularly distinctive for one or another ingressive construction and, consequently, to gain a deeper understanding of the semantic profile of this construction. The article is organized as follows. The next section is devoted to previous studies and presents the main hypothesis and research questions for this article. Section 3 outlines the corpus-based methodology. The quantitative results are presented in section 4 while section 5 contains a discussion of the corpus findings. The conclusions are presented in section 6.

2. PREVIOUS STUDIES

Previous scholarship⁴ has noted differences in the usage patterns of COMA and SMA in relation to the following linguistic factors: event type (2), the semantics of certain adverbs (3), negation (4), and tense (5).

- (2) Le chien de nos voisins [commença/ *se mit] à être sourd. (Peeters, 1993 : 40)
 ‘Our neighbor’s dog [began/*started] to be deaf.’

¹We wish to thank the anonymous referees for their valuable comments which contributed to the overall quality of our text. All errors remain ours.

²Ingressive Aspect is also known as Inchoative Aspect (e.g., Wierenga, 2023) and Inceptive Aspect (e.g., Smith, 1997; Xiao and McEnery, 2004). On the notions of ingressive/inchoative in French linguistics, see Verroens (2018). According to Dik and Hengeveld (1997), ingressivity belongs to a particular subtype of grammatical/viewpoint aspect, namely, phasal aspect distinctions. Phasal aspect operates on lexical/situation aspect (e.g., States, Activities in the sense of Vendler, 1967) in that phasal distinctions divide events up into “phase[s] of development [...] in terms of beginning-continuation-end” (Dik and Hengeveld, 1997: 221). Ingressive aspect focuses on the initial temporal boundary.

³Throughout the article, we have translated COMA as ‘to begin’ and SMA as ‘to start’. We are aware that the inter- and intra-linguistic differences are not straightforward, but we have opted for this consistent translation for the sake of simplicity.

⁴In particular, we refer to enunciative analyses (Franckel, 1989; Sato, 1994; Saunier, 1999), analyses in Natural Semantic Metalanguage (Peeters, 1989, 1993) and logical analyses (Nef, 1980; Gardiès, 1981; Marque-Pucheu, 1999).

- (3) Le soldat amnésique [se met soudain à/ ?commence soudain à] chanter. (Sato, 1994 : 31)
 ‘The amnesic soldier [suddenly starts /? suddenly begins] to sing.’
- (4) Je [n’ai pas commencé/ ? ne me suis pas mis] à manger. (Franckel, 1989 : 144)
 ‘I [haven’t begun /? haven’t started] to eat.’
- (5) Je [commencerais / ?? me mettrais] à travailler. (Sato, 1994 : 32).
 ‘I will [begin / ?? start] to work.’

Several authors (e.g. Lamiroy, 1987; Peeters, 1993; Iordache and Scurtu, 1994; Verroens, 2011) have noted that ingressive verbs hardly take stative verbs as infinitival complements and that for SMA it seems even more difficult than for COMA (2). With regard to the second constraint, it has been noticed that SMA is usually associated with adverbs of velocity and suddenness (e.g. *soudain*, *tout à coup*, *brusquement*, etc. ‘suddenly, abruptly’) (Coseriu, 1976; Peeters, 1993; Saunier, 1999) and that this construction marks a more “brutal” inception than COMA which generates a more “attenuated” inceptive value (Franckel, 1989: 147). Sato (1994) goes so far as to suggest that this type of adverb is exclusively restricted to SMA, in other words that they cannot appear with COMA, as shown in (3). As for the negation expressed in (4), Franckel (1989: 144) judges that the negation would be very strongly constrained, if not even impossible for SMA. In relation to tense, Sato (1994) reports a future constraint (5) for SMA, but not for COMA. According to the latter author, it is very difficult, if not impossible, to have SMA in the future for the simple reason that it would violate the unexpectedness of SMA, while COMA implies anticipation or intentionality.

The corpus-based analysis of Verroens and De Cuyper (2023) shows that most of the intuitive observations illustrated in (2)-(5) are justified. The findings of their mixed-effects logistic regression model suggest that SMA is significantly associated with Activities (in the sense of Vendler, 1967). Furthermore, SMA appears to be associated with the tenses *Passé simple*, *Futur proche* and *Subjonctif présent*, while COMA is associated with *Plus-que-parfait* and *Indicatif imparfait*. As for the adverbs (3) and the somewhat impressionistic assumption of some linguists about the “brutality” of the inception, there is no evidence that this is indeed the case, because of the limited number of instances found with an adverb. In relation to phasal aspect, Verroens and De Cuyper (2023) interpret their results from a frame-semantic perspective (Croft, 2012) and they were able to observe that the two constructions mark a sub-event, namely the initial phase of the event, in a different way: COMA builds a “durative sub-event” corresponding to the initial phase of the event (designated by the infinitive), and this non-punctual initial phase⁵ can just as easily be grasped under a perfective as an imperfective construal⁶ (*commença à/ était en train de commencer à* ‘began at/was beginning at’), whereas SMA rather designates the initial boundary of the event. SMA thus constructs an initial

⁵In line with Selection-Theoretical approaches to aspect (i.a. Bickel, 1997; Michaelis, 2004, 2011; Bogaards, 2022), we use ‘phase’ for durative parts of an event in between temporal boundaries vs. ‘transition’ for punctual starting and endpoints of an event, i.e., its temporal boundaries.

⁶Croft and Cruse (2004) define a construal as a cognitive process by which an experience to be communicated is structured to serve as the semantic representation of a linguistic form or construction.

transition, which, because of its punctuality, is very difficult to reconcile with the imperfective, but is on the other hand perfectly compatible with the perfective and with punctual adverbs. Verroens and De Cuypere (2023) argue that, just like predicates, aspectual constructions can have more than one construal too. The two ingressive constructions can be distinguished on the basis of clear aspectual differences in terms of a punctual/durative analysis. Both ingressive constructions mark the onset, but they modify/coerce⁷ the aspect contour of a base event in different ways. COMA can render the achievement profile or that of the accomplishment while SMA manifests only one profile, more precisely the achievement profile. In other words, COMA, unlike SMA, has the potential to profile the initial phase in two distinct ways. The idea that the two ingressive constructions can be distinguished on the basis of clear aspectual differences fits in with Selection-Theoretical approaches to aspect (i.a. Breu, 1994; Bickel, 1997; Sasse, 2002; Michaelis, 2004, 2011; Croft, 2012; Bogaards, 2022; Koss et al., 2022), which assume that lexical aspect (situation aspect/ Aktionsart) and phasal aspect are built out of the same ingredients, with the latter picking out or coercing the building blocks of the former. The aspectual building blocks shared by lexical and grammatical aspect are temporal boundaries and phases. Selection-theoretical approaches recognize that aspect is not a one-size-fits-all category but is influenced by various linguistic, cognitive, and contextual factors. These approaches aim to uncover the intricate processes through which speakers select specific aspectual forms to convey their intended meanings in different situations.

In this article, we examine the hypothesis of Verroens and De Cuypere (2023) by asking the following questions:

- (i) Which categories of verbs are particularly distinctive for SMA and COMA?
- (ii) Do the results of the collostructional analysis show that both ingressive constructions can be distinguished on the basis of clear aspectual differences in terms of a punctual/durative analysis?

3. DATA AND METHOD

3.1 Data

To check for possible differences according to text type, we used two types of corpora. Our data sample is drawn from the Frantext (FT) literary base for the period 1985 to 2000 and from the journalistic corpus of Le Monde (LM) on

⁷Coercion (i.e. De Swart, 1998; Michaelis, 2004) refers to the process by which a construction that is typically associated with a specific meaning or function is used in a different context, leading to a reinterpretation of its meaning. Coercion often occurs when a construction is extended to cover a broader range of meanings or when it is used in non-canonical contexts. This can result in a construction being “coerced” into a new sense, allowing speakers to convey meanings that might not have been part of the original prototypical meaning of the construction. As such, the event type of the infinitival complement can be an Activity, an Accomplishment, an Achievement, or a State. The event type of the ingressive construction alters that of the infinitival complement and corresponds to an Achievement (e.g. Dowty, 1979: 68, IIIId). According to Verroens and De Cuypere (2023) it can also be an Accomplishment in the case of COMA. From this perspective, SMA has the expected Achievement-profile, whereas COMA has a broader distribution extending beyond Achievements.

CD-ROM (10/2004-9/2006) from which we selected the period January 2005 to September 2006. We collected $N = 2000$ observations: $N = 500$ occurrences per construction (SMA and COMA) per corpus (LM and FT). Note that the number of COMA is greater than that of SMA in the two corpora. For better comparison, we have balanced the corpus, i.e., we have limited it to the same number of occurrences as SMA. If, in Frantext's literary corpus, 4,392,709 words are enough to make 500 tokens, 34,738,595 words are needed in *Le Monde* to obtain the same result. This observation suggests that SMA is less frequent in journalistic texts, as has been noticed before (Roy, 1976: 284; Peeters, 1993: 41–42).

3.2 Method

We have adopted the following procedure: (i) determine which collexemes are typically associated with SMA and COMA, (ii) group these distinctive collexemes into semantic categories, and (iii) establish how this analysis can contribute to the description of the semantic profile of SMA and to the distinction with its quasi-synonym COMA. Below, we explain how the analysis is performed and which labels we assign to the semantic categories.

Collostructional analysis, developed by Stefanowitsch and Gries (2003), is a set of quantitative techniques designed with the aim of measuring the degree of association (collostructional strength) between a slot in a given construction and the collexemes, i.e. the lexical items occupying this slot. Within this approach, it is assumed that “speakers subconsciously perform a statistical analysis of the input and that the statistical associations found in the data are reflected in psychological associations in the mind of the language user” (Stefanowitsch 2006:258). Collostructional analysis has been developed within the framework of Construction Grammar. Consequently, the notion of construction should be understood in the constructional sense (Goldberg 1995, 2006), i.e. the general concept of a form-meaning (or form-function) pair at any level of linguistic structure. What distinguishes collostructional analysis from more traditional collocational analysis is first of all the notion of slot. If collocational analysis takes into account all the words appearing in a determined orbit around the central slot, collostructional analysis is limited to the examination of the words constituting the paradigm of a slot of the construction in question. In comparison to the automatic extraction of lists of co-occurring lexemes characterizing traditional collocational analysis, collostructional analysis also proves to be more adequate and more precise thanks to its statistical basis. The statistical calculations make it possible to measure the degree of association between a collexeme and a construction and it is precisely this sorting between significant and non-significant data that cannot be obtained by a traditional collocational analysis. Collostructional analysis includes different techniques (Stefanowitsch, 2013):

- (i) *Simple collexeme analysis* examines a slot in a given construction, e.g. slot V_{gerund} in the construction [X think nothing of V_{gerund}] (Stefanowitsch and Gries, 2003)
- (ii) *Distinctive collexeme analysis* studies a position in two or more similar constructions, e.g. the verb in the ditransitive and dative prepositional constructions (Gries and Stefanowitsch, 2004a)

- (iii) *Covarying collexeme analysis* looks at the interaction between two positions in a specific construction, e.g. V1 and V2 in the causative construct [X V1 Y into V2_{gerund}] (Gries and Stefanowitsch, 2004b; Stefanowitsch and Gries, 2005)

Distinctive collexeme analysis has mainly been used in the areas of verbal constructions and morphology (e.g. the dative alternation in Gries and Stefanowitsch 2004a; causative constructions in Gilquin, 2006 ; attributive constructions in Lauwers and Van Wettere, 2018), but also to distinguish quasi-synonymous prepositional expressions (Lauwers, 2010). This method is particularly useful for our study since it allows the two ingressive constructions to be separated by identifying the infinitives that are distinctively associated with one or the other.

The collostructional analysis has been conducted with PerlClx 1.0b, a collection of scripts written for Perl by Anatol Stefanowitsch. Like all methods in the collostructional family, it is based on a cross-tabulation of the raw frequencies of the linguistic features and the construction in question. In order to calculate the distinctiveness of a given collexeme, we need four frequencies: the lemma frequency of the collexeme in construction A, the lemma frequency of the collexeme in construction B, and the frequencies of construction A and construction B with words other than the collexeme in question. These can then be entered in a 2-by-2 table (Table 1) and calculated by a Fisher exact test or any other distributional statistic (Gries and Stefanowitsch, 2004a:102).

Table 1. Frequency information needed for a distinctive collexeme analysis

	Construction A	Construction B
lexeme L	Freq (L+A)	Freq (L+B)
-L (other lexemes)	Freq (-L+A)	Freq (-L+B)

In other words, this analysis first requires as input a list with the collexemes of construction A (SMA) and a list with the collexemes of construction B (COMA). This input was acquired on the basis of a manual identification in our dataset where the co-occurring infinitives have been annotated under the label infinitive 1.⁸ The mentioned lists allow the program to determine if a lexeme L appears more frequently in a position of a construction (i.e. SMA or COMA + inf.) than predicted by chance. More precisely, the program first calculates the observed frequency as well as the expected frequency of each collexeme in each construction from a contingency table. Then, the Fisher-Yates exact test determines the collostructional strength by examining whether the frequency of a collexeme with a construct C is distinctive. The threshold value of statistical significance is set at $p < .05$. By multiplying this test for all the collexemes, we obtain a reliable list of all the verbs

⁸Sometimes there were several co-occurring infinitives, but the infinitives 2 and 3 are not part of the actual quantitative analysis. For instance:

(i) Mais, tôt ou tard, il se **mettait** à consulter ses fiches, chercher, fouiner pour satisfaire le client. (FT081)
'But, sooner or later, he would start consulting his files, searching, nosing around to satisfy the customer.'

that appear significantly with one or the other construction. Once the p -values were obtained, we arranged (see Tables 2–4) the collexemes in decreasing order of ‘distinctiveness’ which corresponds to the p -value obtained by the Fisher-Yates exact test. If Stefanowitsch and Gries (2003) consider the p -value as the measure of the collostructional strength, Stefanowitsch and Gries (2005), on the other hand, transform the p -value into a logarithmic value ($=\text{Log}_{10}$) to represent the degree of association. The interpretation is different, while the result remains the same. A high degree of association therefore corresponds to a high Log_{10} value, but also to a very low p -value. Although the repelled collexemes may have a potential interest (Stefanowitsch 2008), we have only retained in the presentation of the results the collexemes which are associated in a significant way with the constructions SMA and COMA.

Once we established the distinctive collexemes per construction we classified them into semantic categories. We rely on the terminology of Levin (1993) to assign the labels of these semantic categories. Categories that turned out to be relevant based on our corpus data are Verbs of Change of Possession (e.g. *to give, to sell*), Change of State Verbs (e.g. *to dry, to become*), Verbs of Communication (e.g. *to yell, to speak*), Conjecture Verbs (e.g. *to know, to recognize*), Verbs of Creation and Transformation (e.g. *to build, to dance*), Declare Verbs (e.g. *to think, to believe*), Verbs of Exerting Force (e.g. *to push, to pull*), Exist Verbs (e.g. *to live, to exist*), Verbs of Motion (e.g. *to run, to turn*), Verbs Involving the Body (e.g. *to smile, to tremble*), Verbs of Perception (e.g. *to see, to feel*), Psych-Verbs (e.g. *to be interested, to worry*), Verbs of Sending and Carrying (e.g. *to send, to take*), and Weather verbs (e.g. *to rain, to snow*).

4. RESULTS

4.1 Results of the literary corpus

First, the results of the literary corpus (Table 2) show that there are as many ($N = 9$) significant collexemes for SMA (N tokens = 500; N types = 249) as COMA (N tokens = 500; N types = 333). Second, the most significant collexemes in relation to SMA are *pleurer* ‘to cry’ and *rire* ‘to laugh’. For these verbs, we note that the difference between the frequency observed in the two constructions is remarkable. Together with *chialer* (‘to blubber, weep noisily’), they all belong to Verbs Involving the Body, more precisely, they can be defined as verbs of non-verbal expression involving facial expressions that are associated with a particular emotion (Levin, 1993:219).

- (8) En entendant le nom de Geoffrey, Jessica et Atalanta se **mirent** à rire toutes les deux. (FT072)
 ‘Hearing Geoffrey’s name, Jessica and Atalanta both started laughing.’
- (9) Je suis allé derrière la baraque et j’ai gerbé mes soixante-dix Néocodion en me demandant ce que je foutais là, je me suis **mis** à chialer, ça m’a fait du bien (FT073)
 ‘I went behind the barracks and I stacked my seventy Néocodion wondering what I was doing there, I started to blubber, it did me good’

Table 2. Distinctive collexemes in *Frantext*

Collexeme	SMA		COMA		Distinctive for	Fisher Yates (ordered)	Log10-FE (ordered)
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency			
pleurer (to cry)	29	16	1	14	SMA	0.0000	6.9482
rire (to laugh)	30	18	3	15	SMA	0.0000	5.5763
courir (to run)	19	11	2	10	SMA	0.0003	3.5532
hurler (to yell)	8	4	0	4	SMA	0.0062	2.2046
crier (to scream)	8	4	0	4	SMA	0.0062	2.2046
chialer (to blubber)	8	4	0	4	SMA	0.0062	2.2046
parler (to talk)	27	20	10	17	SMA	0.0099	2.0038
jouer (to play)	5	3	0	2	SMA	0.0422	1.3746
danser (to dance)	5	3	0	2	SMA	0.0422	1.3746
comprendre (to understand)	3	11	18	10	COMA	0.0002	3.6354
sentir (to feel)	2	9	14	7	COMA	0.0008	3.0738
connaître (to know)	0	4	8	4	COMA	0.0022	2.6535
savoir (to know)	0	3	6	3	COMA	0.0103	1.9868
prendre (to take)	0	3	5	2	COMA	0.0222	1.6543
voir (to see)	0	3	5	2	COMA	0.0222	1.6543
devenir (to become)	3	6	9	6	COMA	0.0456	1.3408
sécher (to dry)	0	2	4	2	COMA	0.0476	1.3223
gémir (to groan)	0	2	4	2	COMA	0.0476	1.3223

Third, the significant collexemes of the two ingressive constructions belong to clearly distinct semantic fields: the distinctive collexemes for SMA all refer to activities, and this in a fairly homogeneous way. More specifically, they refer to acts of non-verbal expression (*pleurer* ‘to cry’, *rire* ‘to laugh’, *chialer* ‘to blubber’), communication (*crier* ‘to scream’, *hurler* ‘to yell’, *parler* ‘to speak’), acts of performance (*danser* ‘to dance’, *jouer* ‘to play’) and motion (*courir* ‘to run’). Generally speaking, we can group them together in the supercategory of ((non-)verbal) expressive collexemes.

The list of distinctive collexemes for COMA is more heterogeneous and thus it is more difficult to formulate a supercategory that could encompass the various subcategories. The distinctive collexemes refer to cognitive acts (*comprendre* ‘to understand, *connaître* ‘to know’, *savoir* ‘to know’), sensory acts (= perception verbs *sentir* ‘to smell’, *voir* ‘to see’), acts of non-verbal expression (*gémir* ‘to groan’), acts of sending and carrying (*prendre* ‘to take’), and change of state (*devenir* ‘to become’, *sécher* ‘to dry’). When we look at the largest group of distinctive collexemes for COMA, the “cognitive collexemes” (*comprendre*, *connaître*, *savoir*) exemplified in (10)–(12), we observe that the COMA construction coerces the basic state or achievement event. The process of knowing becomes more gradual, e.g. in (12) it can be paraphrased by “to become familiar with”. The same can be said for (13), where the perception verb *voir* behaves more like a cognitive verb (= to understand) and in which a gradual process can also be distinguished.⁹

- (10) Il me semble, murmura A de la voix la plus douce, que je **commence** à comprendre, grâce à Chateaubriand et à toi, comment fonctionnent les hommes. (FT075)
 ‘It seems to me, murmured A in the softest voice, that I am beginning to understand, thanks to Chateaubriand and to you, how men function.’
- (11) « Je **commence** à connaître les plantes de la taïga par cœur », dit Albertine, en versant de cette soupe dans leurs assiettes. (FT076)
 ‘I’m beginning to know the plants of the taiga by heart,’ said Albertine, pouring this soup on their plates.’
- (12) A me regarda de ce regard que je **commençais** à connaître et qui ne me voulait pas de bien. (FT077)
 ‘A looked at me with that look that I was beginning to know and that didn’t mean any good to me.’
- (13) Je **commence** à bien voir les grandes lignes. (FT080)
 ‘I’m beginning to see the main lines well.’

⁹Recall that cognitive verbs (*understand*, *know*, *believe*) and verbs of perception (*see*, *hear*, *perceive*) are two-faced in that they have both Achievement-readings and State-readings (i.e. Dowty, 1979:66–68): (i) in case of Achievement-reading, a preparatory phase is added and this results in an Accomplishment; (ii) in case of State-reading, dynamicity/scalarity is added and the state is exhibited to a higher and higher degree, resulting in an Activity. It isn’t evident which of these is targeted by COMA, but, more standardly, cognitive verbs are interpreted as States and perception verbs as Achievements (i.e. Rothstein, 2004). As such, we could state that (12) aligns more with the State-reading (“become [more and more] familiar with”) and (13) with the Achievement-reading.

4.2 Results of the journalistic corpus

A first observation is that SMA (N tokens = 500; N types = 271) has more significant collexemes than COMA (N tokens = 500; N types = 334), i.e. 10 versus 8, the first four of which also appear in the literary corpus. For the journalistic corpus, likewise, the (non-)verbal expressive collexemes are significantly associated with SMA.

- (14) Il n'arrivait pas à parler en public d'Auschwitz, se **mettait** vite à pleurer. Alors il allait aux commémorations avec son « habit de déporté ». (LM083)
 'He was unable to speak in public about Auschwitz, quickly started to cry. So he went to the commemorations with his "deportee's clothes".'
- (15) Un type s'est approché, il s'est **mis** à hurler en arabe, a chargé son arme et s'est mis à tirer. (LM084)
 'A guy approached, he started screaming in Arabic, loaded his gun and started shooting.'

Table 3. Distinctive collexemes in *Le Monde*

Collexeme	SMA		COMA		Distinctive for	Fisher Yates (ordered)	Log10-FE (ordered)
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency			
pleurer (to cry)	17	9	0	8	SMA	0.00002	4.67363
parler (to speak)	25	16	4	13	SMA	0.00018	3.75030
courir (to run)	10	5	0	5	SMA	0.00184	2.73403
hurler (to yell)	8	4	0	4	SMA	0.00655	2.18380
pleuvoir (to rain)	8	4	0	4	SMA	0.00655	2.18380
douter (to doubt)	10	6	1	5	SMA	0.01052	1.97812
vivre (to live)	6	3	0	3	SMA	0.02316	1.63528
trembler (to tremble)	6	3	0	3	SMA	0.02316	1.63528
bouger (to move)	6	3	0	3	SMA	0.02316	1.63528
pousser (to push)	5	3	0	2	SMA	0.04348	1.36167
travailler (to work)	2	7	12	7	COMA	0.00284	2.54668
donner (to give)	0	3	5	2	COMA	0.02142	1.66927

(Continued)

Table 3. (Continued.)

Collexeme	SMA		COMA		Distinctive for	Fisher Yates (ordered)	Log10-FE (ordered)
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency			
constituer (to constitute)	0	3	5	2	COMA	0.02142	1.66927
avoir (to have)	0	3	5	2	COMA	0.02142	1.66927
réfléchir (to think)	1	4	6	3	COMA	0.04186	1.37818
prendre forme (to take shape)	0	2	4	2	COMA	0.04631	1.33429
discuter (to discuss)	0	2	4	2	COMA	0.04631	1.33429
bâtir (to build)	0	2	4	2	COMA	0.04631	1.33429

Verbs which were unattested or repelled in the literary corpus belong to the fields of motion (*bouger* ‘to move’), verbs of exerting force (*pousser* ‘to push’), weather verbs (*pleuvoir* ‘to rain’), psych-verbs (*douter* ‘to doubt’) and verbs of existence (*vivre* ‘to live’). As for COMA, on the one hand, we observe a predominant presence of the collexeme *travailler* ‘to work’, on the other hand, it concerns very diverse semantic fields, namely cognition (*réfléchir* ‘to think’), possession (*avoir*¹⁰ ‘to have’), transfer (*donner* ‘to give’), creation (*bâtir* ‘to build’, *constituer* ‘to constitute’, *prendre forme* ‘to take shape’), and communication (*discuter* ‘to discuss’). In general, we can say that the significant collexemes for COMA in the journalistic corpus belong to such diverse semantic fields that it is no longer appropriate to propose a common denominator.

4.3 Combined results

When we take the two corpora together, we distinguish, out of a total of 1000 tokens of each construction, 14 significant collexemes for SMA (N types = 445) versus 21 for COMA (N types = 592). The strong association between SMA and the collexemes *pleurer*, *rire*, *courir*, *parler* and *hurler* is again remarkable.

Table 4 shows that the observed frequency of these collexemes diverges considerably from one construction to another. As for the verbs with SMA that are not listed for the individual corpora, we note the motion verb *tourner* ‘to turn’ and the cognitive verb *penser* ‘to think’. With regard to the collexemes significant for the COMA construction, they generally refer to more diverse semantic fields like cognitive acts (*comprendre* ‘to understand’, *connaître* ‘to know’, *savoir* ‘to know’), psych-verbs (*s’inquiéter* ‘to worry’, *s’intéresser* ‘to be interested’), sensory acts (*sentir* ‘to feel’, *voir* ‘to see’, *toucher* ‘to touch’), acts of creation (*bâtir* ‘to build’, *constituer* ‘to

¹⁰The collexeme *avoir* covers all attestations of transitive use as in *Maintenant, elle commence à avoir une vraie intelligence de jeu* (LM086) ‘Now she begins to have real game intelligence’, while expressions like *avoir peur* ‘to be afraid’ were annotated separately.

Table 4. Distinctive collexemes in *Frantext* and *Le Monde*

Collexeme	SMA		COMA		Distinctive for	Fisher Yates (ordered)	Log10-FE (ordered)
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency			
pleurer (to cry)	46	25	1	22	SMA	0,0000	11,4163
courir (to run)	29	17	2	14	SMA	0,0000	5,9505
rire (to laugh)	32	19	3	16	SMA	0,0000	5,9504
parler (to speak)	52	35	14	31	SMA	0,0000	4,9296
hurler (to yell)	16	9	0	7	SMA	0,0000	4,3922
crier (to scream)	11	6	0	5	SMA	0,0010	3,0137
pleuvoir (to rain)	11	6	0	5	SMA	0,0010	3,0137
chialer (to blubber)	9	5	0	4	SMA	0,0034	2,4639
danser (to dance)	10	6	1	5	SMA	0,0104	1,9834
penser (to think)	7	4	0	3	SMA	0,0122	1,9148
trembler (to tremble)	7	4	0	3	SMA	0,0122	1,9148
bouger (to move)	6	3	0	3	SMA	0,0229	1,6406
pousser (to push)	6	3	0	3	SMA	0,0229	1,6406
tourner (to turn)	15	11	5	9	SMA	0,0399	1,3992
comprendre (to understand)	3	12	19	10	COMA	0,0001	3,9000
sentir (to feel)	3	10	15	8	COMA	0,0015	2,8212
prendre (to take)	2	7	12	7	COMA	0,0031	2,5155
connaître (to know)	1	5	9	5	COMA	0,0060	2,2246
travailler (to work)	3	8	12	7	COMA	0,0088	2,0559
savoir (to know)	0	3	6	3	COMA	0,0102	1,9907

(Continued)

Table 4. (Continued.)

Collexeme	SMA		COMA		Distinctive for	Fisher Yates (ordered)	Log10-FE (ordered)
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency			
voir (to see)	1	5	8	4	COMA	0,0117	1,9334
s'intéresser (be interested)	2	6	9	5	COMA	0,0192	1,7168
sortir (to go out)	0	3	5	2	COMA	0,0220	1,6583
constituer (to form)	0	3	5	2	COMA	0,0220	1,6583
bâtir (to build)	0	3	5	2	COMA	0,0220	1,6583
s'inquiéter (to worry)	0	3	5	2	COMA	0,0220	1,6583
donner (to give)	1	4	7	4	COMA	0,0225	1,6469
devenir (to become)	3	6	9	6	COMA	0,0452	1,3450
toucher (to touch)	0	2	4	2	COMA	0,0472	1,3261
gémir (to groan)	0	2	4	2	COMA	0,0472	1,3261
prendre forme (to take shape)	0	2	4	2	COMA	0,0472	1,3261
mener (to lead)	0	2	4	2	COMA	0,0472	1,3261
se sentir (to feel)	0	2	4	2	COMA	0,0472	1,3261
sécher (to dry)	0	2	4	2	COMA	0,0472	1,3261
poser (to put)	0	2	4	2	COMA	0,0472	1,3261

constitute'), verbs of sending and carrying (*prendre* 'to take'), change of state (*devenir* 'to become', *sécher* 'to dry'), etc. The data shows that verbs of non-verbal expression (e.g. *rire* 'to laugh', *ricaner* 'to sneer', *sangloter* 'to sob') can also be combined with the COMA construction, but without any significant association.

5. DISCUSSION

The identification of the different collexemes and distinctive semantic domains for both ingressive constructions enables us to examine how our analysis can contribute to the overall description of their semantic profile. Any infinitive can occupy the collexeme position provided that its meaning is semantically compatible with the meaning of the construction or, more precisely, with the meaning assigned by

the construction to the particular slot in which the word appears, i.e. the infinitive slot (Stefanowitsch and Gries, 2003: 213). According to the hypothesis of Verroens and De Cuypere (2023), the ingressive constructions have their own meaning: Both ingressive constructions mark the onset of an event, but they modify the aspectual contour of a base event in different ways. COMA is able to exhibit an achievement profile, an accomplishment profile, or an activity¹¹ profile while SMA tends to mark a punctual transition, i.e. exhibits only an achievement profile. The results from their frame-semantic analysis can now be reinterpreted in the light of our collostructional analysis. The achievement profile manifested by SMA highlights the unexpected and unintentional meaning of their privileged collexemes, viz. the expressive collexemes. In expressions like *se mettre à rire* (~burst out laughing) the beginning of the event is punctual because it is the construction that imposes that meaning on the collexeme. This punctuality is particularly compatible with collexemes that do not presuppose an a priori, i.e. that do not manifest premeditation or intentionality like verbs of non-verbal expression (e.g. *rire* ‘to laugh’, *pleurer* ‘to cry’, *chialer* ‘to blubber’). A similar point has been made by Bogaards (2022) for Dutch ingressives. In his corpus study, Bogaards (2022) reports on special ‘punctual’ lexical ingressive expressions like *in huilen/lachen uitbarsten* ‘burst into crying/laughing’. He observes that “The punctual semantics of ‘bursting’ [...] appears to map to the initial boundary of these situations. This might be facilitated by the fact that the initiation of laughter and crying is usually accompanied by some vehemence” (Bogaards, 2022:17). The infinitive receives a meaning that it does not initially have and which comes from the meaning of the construction which influences the lexemes. On the other hand, COMA is more neutral because it can have several profiles. COMA construes a more gradual beginning of the event, which is illustrated with cognitive collexemes: *commencer à comprendre/savoir* (‘beginning to understand/know’) implies a more gradual beginning of the event. To obtain the same effect with SMA, it is necessary, for example, to introduce the adverb *lentement* ‘slowly’ (16). Due to the specific meaning of SMA, it is not surprising that the combination with *lentement* is rare.¹² Without *lentement*, we fall back again on the usual intrinsic punctual value of the SMA construction (17). The gradual beginning of the event characterizes COMA and even more when it is preceded by an opinion verb like *il me semble que* ‘it seems to me that’ (18), *je crois que* ‘I believe that’ (19), etc. As for the sensory collexemes, SMA only appears next to *sentir* ‘to smell’ (20) in our corpus, while COMA appears in thirteen of the fifteen examples next to *sentir* ‘to feel’ (21). It seems to us that olfactory *sentir* is more compatible with the punctual sense of SMA, while *sentir* in the sense of ‘to feel’ rather requires a more gradual onset of the event. Example (22)

¹¹When COMA targets the State-reading in cases like *commencer à connaître* ‘begin to become (more and more) familiar’.

¹²A reviewer notes that a very similar claim has been made by Van Pottelberge (2004:41–42) about the Dutch ingressive *aan het*-construction with *-slaan* ‘to hit’. Van Pottelberge calls the meaning contribution of *slaan* (in contrast to *gaan* ‘to go’) “schnell, plötzlich, energisch” ‘fast, sudden, energetic’ and this construction would be hardly compatible with ‘slowly’. The existence of this contrast in Dutch suggests that it may be a more general crosslinguistic phenomenon.

illustrates very well the overall meaning of the two constructions as well as their significant collexemes.

- (16) Il se **mettait** lentement à comprendre qu'à un certain niveau de la finance et de la politique américaine les juifs, si extraordinairement commodes par leur agilité intellectuelle dans les tâches (FT045)
 'He slowly started to understand that at a certain level of American finance and politics the Jews, so extraordinarily convenient by their intellectual agility in the tasks'
- (17) La mère aussi se **met** à comprendre combien son enfant est intelligent dans ses réactions. (FT079)
 'The mother also starts to understand how intelligent her child is in his reactions.'
- (18) Il me semble, oui, il me semble que je **commence** à comprendre.(FT083)
 'It seems to me, yes, it seems to me that I am beginning to understand.'
- (19) Moi aussi, hélas ! je crois que je **commence** à comprendre...(FT084)
 'Me too, alas! I think I'm beginning to understand...'
- (20) D'abord, il y eut l'odeur. Un jour, les préservatifs ougandais se sont **mis** à sentir mauvais. (LM087)
 'First, there was the smell. One day Ugandan condoms started to smell bad.'
- (21) à têtes de griffons, entre les deux fenêtres, face aux bustes et aux têtes grecques et romaines, en marbre et en bronze, je me répétais les quelques mots de mon rôle, **commençant** à sentir monter en moi le trac bien connu. (FT085)
 'with the heads of griffins, between the two windows, facing the Greek and Roman busts and heads, in marble and bronze, I repeated to myself the few words of my role, beginning to feel the well-known stage fright rising within me.'
- (22) Il m'a regardé attentivement puis il s'est **mis** à sourire. Je restais méfiant mais je **commençais** à me sentir mieux, il avait l'air pas mal ce type, j'étais peut-être tombé sur un bon numéro pour une fois.(FT086)
 'He looked at me attentively then he started to smile. I remained wary but I was beginning to feel better, he looked pretty good, this guy, maybe I had come across a good number for once.'

The collostructional analysis clearly demonstrates the inherent meaning of both constructions. The incremental value of COMA is compatible with the gradual onset of cognitive and sensory collexemes.

On the other hand, there is the punctual meaning of the SMA construction, which is specific to it, i.e. it is not inferred by the collexemes. The punctual value of SMA is not only compatible with expressive collexemes, but, moreover, highlights their unforeseen and unintentional meaning. We can identify a clear aspectual distinction in terms of punctual (SMA) vs. durative (COMA) analysis, which is in line with the analysis of Verroens and De Cuypere (2023), i.e. both ingressives are able to alter a basic event, but COMA can render the achievement profile, the accomplishment profile, or the activity profile while SMA manifests only one profile, more precisely the achievement profile.

6. CONCLUSION

This article has examined the semantic value of the infinitive in the ingressive constructions SMA and COMA using distinctive collexeme analysis. This method makes it possible to distinguish quasi-synonymous constructions by identifying which collexemes are typical of one or the other construction. The results of the two types of corpora, the literary Frantext corpus and the journalistic corpus of *Le Monde*, are quite similar. In general, we find that there were several collexemes which are strongly linked to the constructions SMA and COMA. The significant collexemes that come into play for SMA are essentially part of the semantic classes of non-verbal (*crying, laughing, whining*) or verbal (*shouting, yelling, speaking*) expression, acts of performance (*dancing, playing*), verbs of exerting force (*pushing*), and motion (*running, moving, turning*). The collexemes significant for the construction SMA are fairly homogeneous across the different corpora and we can group them into the general category of expressive collexemes. On the other hand, the collexemes significant for COMA are more heterogeneous and belong, in addition to the category of cognitive collexemes (*understanding, knowing*), also to the semantic fields of sensory (*feeling, seeing*) and creative (*building*) acts. The results are compatible with the hypothesis put forward by Verroens and De Cuyper (2023) stating that the overall meaning of the SMA construction is intrinsically punctual, i.e. is not inferred by collexemes. The punctual value of SMA is not only compatible with verbs of (non-)verbal expression, but, moreover, emphasizes their unforeseen and unintentional meaning. Conversely, the incremental value of COMA is consistent with the gradual onset of cognitive and sensory collexemes. Finally, a perspective for future research could be a global study including also the much rarer ingressive constructions *partir à, se foutre à, and se prendre à* in order to establish the similarities and differences with the semantic profile of SMA and COMA. For the time being, we consider COMA as the prototypical construction in the ingressive construction because of the transparent meaning of the verb *commencer* ('begin'), fewer distributional constraints (e.g. more collexeme types) and the ability to have more than one construal, i.e. an achievement, accomplishment or activity profile.

Competing interests. The author(s) declare none.

References

- Bickel, B. (1997). Aspectual scope and the difference between logical and semantic representation. *Lingua*, 102(1-2): 115–131.
- Bogaards, M. (2022). The discovery of aspect: A heuristic parallel corpus study of ingressive, continuative and resumptive viewpoint aspect. *Languages*, 7:158.
- Breu, W. (1994). Interactions between lexical, temporal and aspectual meanings. *Studies in Language*, 18:23–44.
- Coseriu, E. (1976). *Das romanische Verbalsystem*. Tübingen: TBL Verlag Gunter Narr.
- Croft, W. (2012). *Verbs : aspect and causal structure*. Oxford: Oxford University Press.
- Croft, W., and Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dik, S. C., and Hengeveld, K. (1997). *The Theory of functional grammar*. Part 1: The structure of the clause (2nd, rev. ed.). Berlin-New York: Mouton de Gruyter.
- Dowty, D. R. (1979). *Word meaning and Montague grammar : the semantics of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht: Reidel.

- Franckel, J.-J.** (1989). *Étude de quelques marqueurs aspectuels du français*, Langue et cultures 21. Genève: Droz.
- Gardiès, J.-L.** (1981). Éléments pour une grammaire de l'aspect. *Modèles linguistiques*, 3:112–134.
- Gilquin, G.** (2006). The verb slot in causative constructions. Finding the best fit. *Constructions*, SV1-3/2006. [www.constructions-online.de,urn:nbn:de:0009-4-6741].
- Goldberg, A.** (1995). *Constructions : A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A.** (2006). *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.
- Gries, S. T., and Stefanowitsch, A.** (2004a). Extending colostruational analysis. *International Journal of Corpus Linguistics*, 9(1):97–129.
- Gries, S. T., and Stefanowitsch, A.** (2004b). Covarying collexemes in the *into*-causative. In M. Achard and S. Kemmer (Eds.), *Language, Culture and Mind*. Stanford CA: CSLI, pp. 225–236.
- Iordache, E.-R., and Scurtu, G.** (1994). Etude sémantique et syntaxique des périphrases verbales marquant le début d'accomplissement du procès. *Cahiers de Linguistique Théorique et Appliquée*, XXXI: 41–48.
- Koss, T., De Wit, A., and van der Auwera, J.** (2022). The aspectual meaning of non-aspectual constructions. *Languages*, 7 (143). doi:[10.3390/languages7020143](https://doi.org/10.3390/languages7020143)
- Lamiroy, B.** (1987). The complementation of aspectual verbs in French. *Language*, 63: 278–298.
- Lauwers, P.** (2010). Comment dissocier des locutions prépositives quasi-synonymiques? Essai d'analyse colostruactionnelle. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 55(1): 55–84.
- Lauwers, P., and Van Wetteer, N.** (2018). *Virer et tourner* attributifs: De l'analyse quantitative des cooccurrences aux contrastes sémantiques. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 63(3): 386–422.
- Levin, B.** (1993). *English verb classes and alternations: a preliminary investigation*. Chicago (Ill.): University of Chicago Press.
- Marque-Pucheu, C.** (1999). L'inchoatif : marques formelles et lexicales et interprétation logique. In S. Vogeeler (Ed.), *La modalité sous tous ses aspects*. Amsterdam: Rodopi, pp. 233–257.
- Michaelis, L.** (2004). Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics*, 15(1): 1–67.
- Michaelis, L.** (2011). Stative by construction. *Linguistics*, 49(6), 1359–1399.
- Nef, F.** (1980). Les verbes aspectuels du français: remarques sémantiques et esquisse d'un traitement formel. *Semantikos*, 1 : 11–46.
- Peeters, B.** (1989). *Commencement, continuation, cessation : a conceptual analysis of a set of English and French verbs from an axiological point of view*. Doctoral dissertation. Australian National University, Canberra.
- Peeters, B.** (1993). *Commencer et se mettre à : une description axiologico - conceptuelle*. *Langue française*, 98 : 24–47.
- Rothstein, S.** (2004). *Structuring events : a study in the semantics of lexical aspect*. Malden (Mass.): Blackwell.
- Roy, G.-R.** (1976). *Contribution à l'analyse du syntagme verbal : étude morpho-syntaxique et statistique des coverbes*. Québec: Presses de l'Université de Laval.
- Sasse, H.-J.** (2002). Recent activity in the theory of aspect: accomplishments, achievements, or just non-progressive state? *Linguistic Typology*, 6: 199–271.
- Sato, J.** (1994). Valeurs sémantiques de *se mettre à* et *commencer à*. *Bulletin d'études de Linguistique française*, 28: 30–35.
- Saunier, E.** (1999). Contribution à une étude de l'inchoation : « se mettre à + inf. ». Contraintes d'emploi, effets de sens et propriétés du verbe *mettre*. In S. Vogeeler, et alii (Ed.), *La modalité sous tous ses aspects*. Amsterdam: Rodopi, pp. 259–288.
- Smith, C. S.** (1991 [1997]). *The parameter of aspect* (2nd (First edition: 1991) ed.). Dordrecht: Kluwer academic.
- Stefanowitsch, A.** (2006). Distinctive collexeme analysis and diachrony: A comment. *Corpus Linguistics and Linguistic Theory*, 2(2): 257–262.
- Stefanowitsch, A.** (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, 9(3): 513–531.
- Stefanowitsch, A.** (2013). Colostruational Analysis. In T. Hoffmann and G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar*. Oxford : Oxford University Press, pp. 290–306.

- Stefanowitsch, A., and Gries, S. T.** (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, **8**(2): 209–243.
- Stefanowitsch, A., and Gries, S. T.** (2005). Covarying Collexemes. *Corpus Linguistics and Linguistic Theory*, **1**(1): 1–46.
- Swart, H. de** (1998). Aspect shift and coercion. *Natural language and linguistic theory*, **16**: 347–385.
- Van Pottelberge, J.** (2004). *Der am-Progressiv: Struktur und parallele Entwicklung in den kontinentalwestgermanischen Sprachen*. Tübingen: Gunter Narr.
- Vendler, Z.** (1967). *Linguistics in philosophy*. Ithaca (N.Y.): Cornell University Press.
- Verroens, F.** (2011). *La construction inchoative se mettre à: syntaxe, sémantique et grammaticalisation*. Doctoral dissertation, Ghent University, Ghent.
- Verroens, F.** (2018). Sur la notion d'inchoatif en linguistique française. *Travaux de linguistique*, **76**: 91–111.
- Verroens, F., and De Cuypere, L.** (2023). French ingressive and (phasal) aspect. A frame-semantic corpus-based analysis. *Canadian Journal of Linguistics/ Revue canadienne de Linguistique*, **68** (3): 435–461.
- Wierenga, R.** (2023). “Gaan loop speel!”: Die inchoatieve niehoofwerkwoorde *gaan* en *loop*. *Tydskrif Vir Geesteswetenskappe*, **63** (2): 346–363.
- Wilmet, M.** (1998). *Grammaire critique du français*. 2e éd. Paris: Hachette.
- Xiao, R., and McEnery, T.** (2004). *Aspect in Mandarin Chinese: A corpus-based study*. Amsterdam: John Benjamins.