Journal of English for Academic Purposes 54 (2021) 101046

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Let's say: Phraseological patterns of SAY in academic ELF communication

Ying Wang [a,*], Henrik Kaatari [b]

[a] *Karlstad University, Sweden*
[b] *University of Gävle, Sweden*

ARTICLE INFO

Keywords:
Formulaicity
Communication verb SAY
Spoken ELF academic discourse
Event type
Discipline
Individual preference

ABSTRACT

Formulaic sequences (e.g., *on the other hand, for example, as can be seen*) are prevalent in academic discourse. Apart from their various functions, research in the field has uncovered a range of contextual and individual factors associated with the use of formulaic sequences, including genre, discipline, and the user's L1 as well as expertise level. However, most previous studies focus on written discourse and employ a frequency-based approach (e.g., lexical bundles, n-grams). The inherent limitations of the approach are of particular relevance to ELF communication, which involves a high degree of flexibility adaptability. The present study aims to explore features of formulaicity in spoken ELF academic discourse. Through a close examination of the phraseological patterns of one verb SAY in a one-million-word corpus of spoken ELF communication in academic settings, the present study is able to overcome some of the limitations of the frequency approach, thereby shedding further light on formulaicity in language use characterising this particular community and its relationship with factors such as event type, discipline, and individual preference.

1. Introduction

Formulaic sequences (FSs), i.e., fixed or semi-fixed word combinations that have "an especially strong relationship with each other in creating their meaning" (Wray, 2008, p. 9), present a major hurdle for L2 learners to achieve nativelike fluency (Pawley & Syder, 1983; Sinclair, 1991; Wray 2008, 2012). Unlike L1 speakers, for whom those 'prefabricated' sequences are stored and retrieved from memory as whole entities at the time of use in accordance with the 'idiom principle' (Sinclair, 1991), L2 speakers are said to rely primarily on 'analytical processing' (Wray, 2002) (or the 'open-choice' principle' in Sinclair's (1991) terminology), building sentences from individual words and grammatical rules, which sometimes results in grammatically correct but non-idiomatic language use. This view has been called into question in studies of English as a Lingua Franca (ELF), which have provided further insights into the working of these two mechanisms in L2 speakers, taking into consideration social and cultural circumstances specific to ELF communication (e.g., House, 2009; Mauranen, 2012; Metsä-Ketelä, 2012; Vetchinnikova, 2015). However, most studies focus on uninterrupted linguistic forms, either fixed combinations such as *you know* (House, 2009), *and so on* (Metsä-Ketelä, 2012) or frequently occurring lexical bundles such as *it is important to, one of the* (Mauranen, 2012; Wang, 2017), whereas sequences that allow a certain degree of formal variability (e.g., *I would/should/will say, as we often say*) have received less attention. As shown in Wang (2018a, 2020), such sequences make up a considerable proportion of FSs produced in ELF settings, revealing important features of formulaicity that call for further

* Corresponding author. Karlstad University: Karlstads Universitet Universitetsgatan, 2 Karlstad 651 88, Sweden.
E-mail address: ying.wang@kau.se (Y. Wang).

https://doi.org/10.1016/j.jeap.2021.101046
Received 2 May 2021; Received in revised form 27 August 2021; Accepted 1 September 2021
Available online 1 September 2021

and more systematic study.

Over the last decade, corpus studies have revealed that FSs are particularly prevalent in academic discourse, providing an important means for, among other things, discourse building (e.g., *on the other hand, as a result, as can be seen*) and stance taking (e.g., *we argue that, to some extent*) (Biber et al., 2004; Cortes, 2004; Hyland, 2008a). Apart from their main functions, research in the field has also uncovered a range of contextual and individual factors tied up with the use of FSs, including genre (e.g., Wang, 2017; Biber et al., 2004), discipline (e.g., Durrant, 2017; Hyland, 2008a), and the language user's L1 as well as expertise level (e.g., Hyland, 2008b; Ädel & Erman, 2012). Most of these studies are based on written academic discourse, but it remains to be seen if and how these factors work together to inform the choice and use of FSs in spoken academic discourse.

The present study takes a step further in the exploration of formulaicity in language use characterising academic ELF communication by focusing on one verb SAY in a large corpus, which allows a systematic investigation of the forms and functions of word combinations involving SAY and their relationship with certain contextual and individual variables.

2. Formulaic language use in academic ELF communication

A range of contextual factors have been observed to be associated with the use of FSs in academic ELF communication. Metsä-Ketelä (2012), for instance, finds that vague expressions (e.g., *and so on, so to say, in a way*) are more commonly used in dialogic speech events such as doctoral defences and seminar discussions than in monologic speech events, and more commonly used in some disciplines such as technology than in humanities and social sciences. The effects of genre (or speech event type) and discipline are also evident in the use of four-word lexical bundles (e.g., *I would like to, I don't know if*) (Wang, 2017). In particular, the latter variable has attracted a great deal of attention in written academic discourse. Using a lexical bundles approach, corpus studies (e.g., Cortes, 2004; Hyland, 2008a, 2008b; Durrant, 2017) have yielded strong evidence that FSs play an important role in identifying membership of a community that shares the same domain of knowledge. Investigations of the effect of this factor on speech have lagged behind, partly due to the hypothesis that being fundamentally different from academic writing, disciplinary differences may be less relevant in academic speech (Ädel, 2008). However, there is some evidence that disciplines also matter in academic speech (e.g., Wang, 2017; Kashiha & Heng, 2014).

Apart from genre and discipline, inter-speaker variation has been observed as regards individual preferences for particular FSs, associated with a speaker's individual and social identity, L1, culture, and personality traits such as extroversion (e.g., Gablasova et al., 2017; Metsä-Ketelä, 2012; Pitzl, 2012). Gablasova et al. (2017), for instance, examine epistemic stance markers (e.g., *I'm certain*) in speech of 132 advanced L2 speakers from different L1 backgrounds in four speaking tasks and find that while the speakers are capable of choosing the most suitable markers according to the speaking task, they also "retain their more individual preferences for particular markers" (p. 626).

One feature of formulaicity in ELF communication that has been well established is, according to Mauranen (2009, 2012), approximation of conventional forms, meaning that there is often considerable variability (or creativity) in linguistic forms employed by ELF speakers. As summarised in Pitzl (2012), the creativity may take the form of lexical substitution (e.g., *I'm not **very** sure* instead of *I'm not **quite** sure*), syntactic and morphosyntactic variation (e.g., *carved in stones* instead of *carved in stone*, ***in** the right track* instead of ***on** the right track*), or internal modification through the insertion of additional elements (e.g., *two **different** sides of the same coin*). Vetchinnikova (2015) brings to light an additional process, namely fixing, through which the core collocates of related entities fuse together; for instance, *there exists research* and *existing research* in her data seem to be derived from the core collocates with *there* and the idea of closing *existing gap* in academic discourse. Despite a certain degree of variability in the surface structure, a holistic unit of meaning or function is still retained, suggesting that the idiom principle is in operation after all. Indeed, as observed in the above studies, formal variations do not seem to disrupt ELF interaction, wherein linguistic creativity is not necessarily a matter of deficiency but rather a conscious choice made to facilitate communication with interactants who have little in common culturally and socio-linguistically. In other words, without shared conventions or formulaic ways of expressing ideas, speakers may intentionally resort to forms that are semantically transparent and syntactically regular (Kecskes, 2019). In addition, the creation of new forms and expressions may also grow out of a need for non-conformity and self-expression, a way to show one's identity (Pitzl, 2012). To put it briefly, these studies show great potential for examining formal variations in formulaic language used by ELF speakers in connection with the two processing mechanisms (analytical vs. holistic) and the relationship between them and relevant contextual factors. The present study takes a step further in that direction through a systematic investigation of all forms of variation associated with one particular word.

As can be seen above, most of the previous studies either focus on a predetermined list of FSs with fixed forms (mostly idioms and metaphors) and a particular function, or rely on the computer to identify chunks of language on the basis of the recurrence of uninterrupted linguistic forms.[1] The latter type of research is probably more helpful in providing an overall picture of formulaicity than does the former. However, the inherent limitations of the methodology mean that a great deal may have been overlooked (Ädel & Erman, 2012; Pan et al., 2020). One limitation is related to dispersion – that is, how bundles are distributed across the whole dataset. It is measured by the number of text files in which a certain word combination occurs, regardless of how often it occurs in each file as long as it reaches the dispersion threshold. However, the high frequency of a given combination could still be skewed by a small number of

[1] Taking the lexical bundles approach a step further, Römer and her colleagues (e.g. Römer, 2010) developed a model to identify p-frames which involve item-internal variation such as *the * and the, by the * of*, where * indicates an open slot that can be filled by any item. While this approach captures internal variation to a certain extent, it is still essentially a frequency-based approach with its inherent limitations.

individuals who, for whatever reason, may have a particular preference for the combination. It can be even more problematic when we deal with spoken corpora where one file can contain multiple speakers. Without taking into account inter-speaker variation in each file, the reliability of previous conclusions on genre and disciplinary differences could be challenged.

Another limitation of the frequency-based approach lies in the difficulty in handling discontinuous and infrequent sequences. The limitation may be particularly relevant to studies dealing with spoken language, where formulaic word strings can be 'interrupted' by hesitations, repetitions, false starts, and the like (Wang, 2018a, 2020). The problem is further compounded in studies of ELF communication, which is characterised by a high degree of flexibility and adaptability, meaning that the language use is constantly adapted by its users according to the context (Seidlhofer, 2011). Formal variations of otherwise formulaic language may escape automatic identification as each variant form may not occur frequently enough to reach the frequency threshold to be included. However, taken together, these forms can reveal important patterns that are crucial in understanding features of formulaicity specific to a given community (Wang, 2018b).

This one-word study is designed specifically to address the limitations of the frequency-based approach in formulaic language research. Focusing on one particular word and its phraseological patterns makes it possible to take into account all possible formal variations, which in turn will allow us to revisit certain issues such as the two processing mechanisms (analytic vs. holistic) in L2 language use as well as the relationship between formulaic language use in spoken academic discourse and extralinguistic factors such as event type, discipline, and speaker variation.

The word SAY emerges from Wang (2020), which is based on a fairly small dataset, as a good candidate for a more systematic investigation, given its high frequency and versatility. In terms of versatility, as Shinzato (2004) notes, speech act verbs such as SAY are productively used in a range of constructions (e.g., *I say, say, I dare say, as you say, that is to say*). In many constructions, the actional meaning of the verb is backgrounded, serving primarily pragmatic functions such as expressing tentativeness or surprise, introducing a condition or an example, and attracting attention (Brinton, 2008). However, pragmatic markers of verbal origins have received much less attention than those of adverbial origin (e.g., *actually*) in the research literature of the English language in general (Van Olmen, 2013), let alone in ELF contexts. While some of the constructions are quite fixed and allow little internal variation such as *that is to say*, others seem formally flexible, involving a number of variable slots including the subject, the form of the verb, the addition of modal verbs and other modifiers. As Swales (2019) suggests, syntactic features such as tense choices may also be an integral part of phraseological patterns and uses, and therefore deserve more attention than they have hitherto received in formulaic language research. While those semantically transparent and syntactically regular expressions may appear to be subject to the open-choice principle, as Vetchinnikova (2015) suggests, they can display 'hidden' formulaic features shared by members of a given discourse community. Through examining the usage patterns of the verb SAY extracted from a one-million-word corpus, the present study keeps digging for what may be 'hidden' from the conventional frequency-based approach. In particular, the following research questions will be addressed.

RQ1. What forms are used and what functions do they have?

RQ2. How are these forms and functions distributed among event types, disciplines, and individual speakers?

3. Data and methods

The present study employed the whole ELFA corpus (Mauranen, 2008), which contains one million words of naturally-occurring spoken academic ELF, transcribed from approximately 131 hours of speech events, both monologic ones (33%), such as lectures and presentations, and dialogic events (67%), such as seminars, thesis defences, and conference discussions, recorded in universities based in Finland. The corpus covers a number of disciplinary domains including social sciences (29%), technology (19%), humanities (17%), natural sciences (13%), and medicine (10%). The data were produced by approximately 650 speakers from 51 different L1 backgrounds (see Mauranen et al., 2010 for more information about the corpus).

AntConc (Anthony, 2020) was used to search for all forms of the verb SAY (*say, says, saying, said*). The concordance lines were examined to retrieve multi-word sequences that form complete semantic units (Buerki, 2016). Although SAY as a single-word unit is also interesting (see, e.g., Brinton, 2008; Van Olmen, 2013), it falls outside the scope of the present study. Therefore, instances such as (1) were excluded from the subsequent analysis.

(1) *can we can we well think of* ***say*** *a country in which or countries in which this process is is happening …* (USEMD280)

In some cases, the search word may be repeated or corrected immediately as in (2) and (3). In (2), the word form *say* occurs twice in the question, but was only counted once in the study. In (3), although *say* was uttered, it was corrected immediately. As the intended expression was *as he writes*, such instances were not considered.

(2) *er can you* ***say*** *s-er perhaps say something* (CDIS08B)

(3) *and as he* ***say****-er writes on page ten where communication technologies provide a public and social space for the constant process on national imagination …* (UDEFD070)

Instances where *say* is used as a noun, as in (4), were also excluded.

(4) *… because i know that the rotumans don't have much* ***say*** *in the fiji daily politics.* (CDIS08A)

The remaining instances were then coded for formal, functional, contextual (event type, discipline), and speaker information. Formal variables include subject, modal verb, modifier, and tense. A functional taxonomy was developed on the basis of dictionary definitions and existing literature such as Brinton (2008), which focuses on comment clauses involving *say*, and Halliday's (2014) metafunctional systems (see also Wang, 2018a,b). The taxonomy, presented in Table 1, covers the main functions of SAY-expressions identified in the present study.

The functions can be roughly divided into three broad categories. The first category (actional) is associated with the core meaning of the communication verb SAY. Expressions of this category are often followed by what is said/written. The other two categories (discoursal and interpersonal) represent pragmatic uses of SAY-expressions. The filler function is not covered in the literature about SAY-expressions, but was added in the present study. It was assigned when a given SAY-expression seems to be dangling outside a proposition; it does not contribute to the meaning of the proposition, and is used to fill the void when the speaker is looking for words or organizing mentally incoming thoughts and ideas. Such expressions are often accompanied by repetitions or hesitation markers.

The concordance line, and sometimes even larger context, was consulted when deciding the function of a given expression, which can have different functions in different contexts (cf. Section 4.3). Some expressions can have multiple functions in the same context, particularly *I would say, I have to say*, where the choice of modal verb alone expresses a kind of attitude (*would*: to express a certain degree of uncertainty; *have to*: to add emphasis). When they are used to introduce an answer or to interrupt, they can be considered as having at least two functions: attitude marker (expressing opinion) and turn-taking device (interactional). In order not to inflate the numbers, each expression was assigned only one function. In cases where it is possible for an expression to have more than one function of the same broad category, the additional function was prioritised over the 'inherent' one. Thus, when expressions such as *I would say* and *I have to say* are used to elicit an answer or a turn, as in (5), they were treated as interactional devices.

(5) <S2> … *how do you know that the extractions you made are adequate* </S2>
<S1> *well* ***i have to say that*** *er er without the modelling of the data values it's not entirely appropriate yet er …* </S1> (UDEFD060)

In cases where there are competing interpretations involving functions belonging to different broad categories, the expression was put under an additional category (Indeterminate). What follows is an example of this kind.

(6) *and i thought this gentleman from india was er looking very obscure and so on you know about that information and so on and @i was like (anyway)@ and and you know i don't know* ***i don't have anything to say*** *but i'm just very excited that i would have almost missed this presentation going through could be so obscure and (xx) and actually it is first-class (critique)* (CDIS08B)

In (6), the expression *I don't have anything to say* could mean that the speaker is not sure or willing to make any comment on the issue, in which sense it would be regarded as expressing opinion (interpersonal). Another possible interpretation is that this expression is simply a filler (discoursal); together with the repetition (*and and*) and hesitation markers (*you know*, *I don't know*), it fills a pause in the ongoing turn for the speaker to retain the floor while thinking about what to say next.

4. Results and discussion

In this section, the main results will be presented first in terms of the overall distribution of SAY-expressions across event types, disciplines, and individual speakers (4.1), and their main forms and functions (4.2). Section 4.3 then discusses some interesting features that have emerged from the results in more detail.

4.1. Overall distribution of SAY-expressions

Altogether, 3539 SAY-expressions were identified in 153 (out of 161) files, produced by 413 different speakers. Fig. 1 shows the distribution and density of SAY-expressions across individual speakers (median (4.0) marked by the vertical line, and the mean (8.57) marked by the red circle; outliers marked by smaller black dots (SD = 13.517)).

The majority of speakers (279 out of 413, 68%) produced one to six SAY-expressions each, while 44 (11%) speakers contributed between 20 and 52 instances each. As can be seen in Fig. 1, four points lie far from the rest, representing two speakers from International Relations with 83 instances each, one from Information Technology (86) and one from Philosophy (139).

Fig. 2 displays the distribution of SAY-expressions across event types, grouped into two broad categories (dialogic vs. monologic). Given the communicative nature of the word, it is not surprising to see SAY-expressions occurring more frequently in dialogic events that involve discussions (3.65 per 1000 words) than in monologic events such as lectures and presentations (2.56 per 1000 words). The difference between dialogic and monologic events is statistically significant: Log-likelihood = 89.72; $p < 0.0001$; odds ratio = 1.43.

Fig. 3 presents the distribution of SAY-expressions across disciplines. Apart from engineering and medicine disciplines, most of which appear to cluster at the two ends of the continuum,[2] it is difficult to draw any conclusion about disciplines belonging to the domains of social sciences and humanities as they are scattered along the whole continuum. Taking event types into consideration, the data of the top five disciplines are dominated by dialogic events such as conferences and doctoral defense discussions and the bottom

[2] Some of the disciplines from these two domains, such as Automatic Engineering, Biology, and Internal Medicine are also scattered along the continuum.

Table 1
Classification of functions.

Broad category	Function + Explanation	Example
Actional	**Speaking** (+speech): to say something to somebody, using words. In the present study, the verb with this function is often followed by a quotation from a source or a direct or indirect speech act.	- *in this book **he says** <* READING ALOUD *> i have no doubt that president urho kekkonen of finland collaborated closely and personally with the kremlin </*READING ALOUD*>* (USEMP01A) - ***they say** in germany* <FOREIGN> *molch* </FOREIGN> (CPRE03E)
	Giving written information: to convey specified information, normally associated with an inanimate subject.	- *then of course there are **labels that** people do trust **say something like** fair trade* (CDIS08B)
Discoursal	**Structuring signal**: to organise long stretches of discourse such as presentations, often together with adverbials such as *first, to conclude* to indicate topic shift.	- *but for my presentation i have chosen the political issue and especially the development and evaluation of political parties so first **i would like to say about** er evaluation of political system at a whole as a whole er …* (USEMP11A)
	Reference: to refer to earlier utterances.	- *and **as i said earlier** it's very much a home-grown method* (CDIS050) - *well er i think erm what just er going back to **what you just sa-said** <*S1*> [mhm] </*S1*> [**before**] indeed it would be nice er it would be a nice er way of a nice kind of research would be to see whether …* (USEMD280)
	Exposition: to elaborate on an element in the discourse by restatement or re-presentation.	- *[this set] and this set of rows erm constitutes the whole (VN) so **that is to say** that this constitutes the whole domain of the function that we want to separate* (UDEFD120)
	Exemplification: to elaborate on an element in the discourse by giving examples.	- *you model a person and you model **let's say** a policeman as er a person belonging to police department …* (UDEFD040)
	Filler: to fill pauses in an utterance when the speaker is looking for a word, planning the discourse, etc.	- *er they are like chains chains that you can't skip one **how do i say** one joint you can't skip one joint they are like chain lined up …* (ULECD030) - *for this reason er you have different er possible ways to express things i think this er **let's say** er differentiation would make your hybrid languages or er <*S1*> mhm-hm </*S1*> classification er* (UDEFD040)
Interpersonal	**Expressing opinion**: to state an opinion or attitude with a certain degree of (un)certainty.	- *and then er of course er with regard to the other influence factors er that is that is a general problem **i would say*** (UDEFD020)
	Interactional device: to elicit comments or to indicate turn-taking[1] (e.g., interrupting; introducing an answer to the question raised).	- *please go ahead **just say something*** (ULECD030) - *what else what else you have if you are in the situation of dystopia **does anyone wanna say something** …* (ULECD030) - *<*S3*> … but it is so crucial for the future that you and the firm you are working with cannot bypass this question as we ar- idea was rising to the surface in your paper maybe it's now on the agenda</*S3*>* *<*S2*>er, **let's say that** we are working on it and … </*S2*>* (UDEFD030)

[1] The position where the SAY-expression occurs (namely, at the beginning of a turn in a conversation) was relied upon when assigning this function.

five by monologic events such as lectures and seminar presentations, suggesting that the distribution is more likely to be affected by event types than disciplines. It should also be mentioned that the contribution from the top four outliers as seen in Fig. 1 does not make the disciplines to which they belong (Philosophy, International Relations, Information Technology) stand out in comparison and therefore the distribution is unlikely to be skewed much by individual differences, either.

4.2. Forms and functions

Table 2 presents the counts (in terms of both tokens and types) of SAY-expressions involving different forms of the verb. As can be seen, the SAY-expressions identified from the dataset are dominated by two forms: *say* and *said*.

Table 3 lists the most common usage patterns for each form. Apart from *let's say* and *so to say*, which can be regarded as fixed expressions in terms of their formal invariability, all the others seem to be semantically and syntactically regular expressions.

It is also clear from Table 3 that different forms prefer different usage patterns, which in turn are associated with different functions. The form *say* is predominantly used with modal or semi-modal verbs (*would, can, must, might, could, have to*) and first- and second-person pronouns (*I, we, you*) to express an opinion or to start a turn in a conversation. *Said* is used primarily as a past tense form, and often co-occurs with the conjunction *as* in the pattern *as somebody said* to refer to what has been said previously. The form *says* is used predominantly with *he* or *it* or a personal name to introduce a quote or (in)direct speech. The *-ing* form of *say* is mostly used in

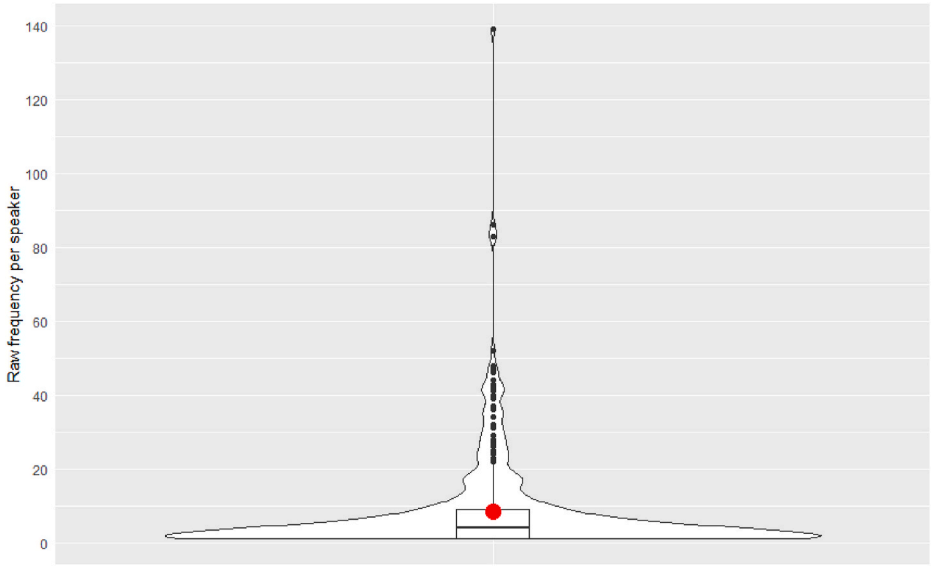

Fig. 1. Distribution of *say*-expressions among speakers.

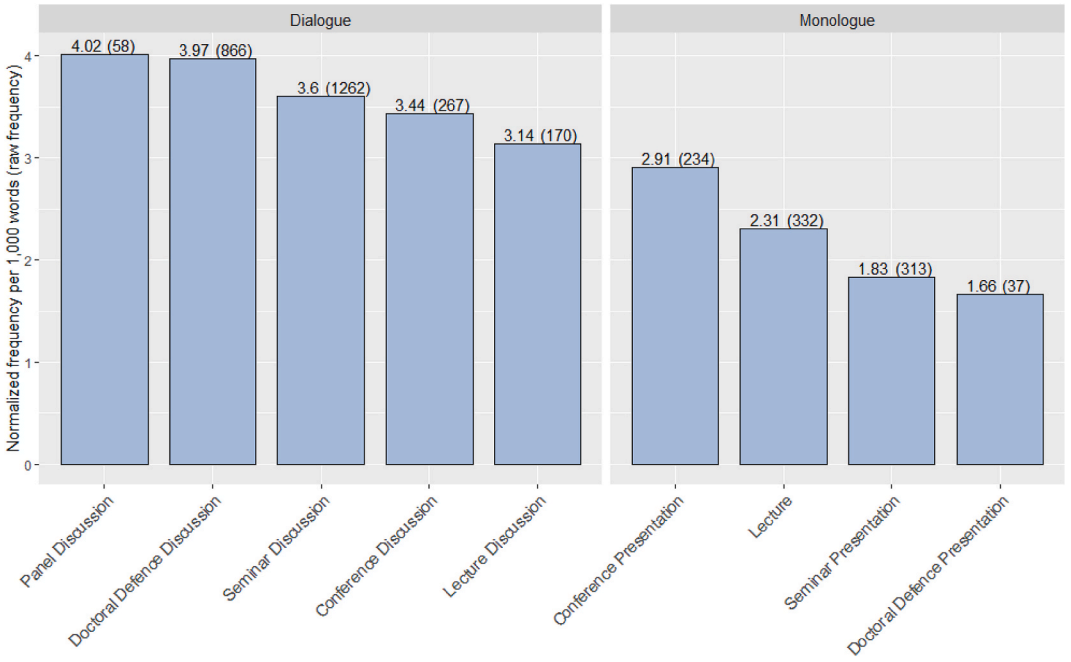

Fig. 2. Distribution of SAY-expressions across event types.

association with the progressive aspect as a way to elaborate further on a point just made.

Fig. 4 presents the distribution of functions according to the three broad categories across dialogic and monologic event types.

Log-likelihood tests of the difference between dialogic and monologic event types within each function type have been conducted. All tests show a statistically significant difference, but with a small to medium effect size (apart from 'indeterminate'): Discoursal: Log-likelihood = 17.69; $p < 0.0001$; odds ratio = 1.28. Interpersonal: Log-likelihood = 39.03; $p < 0.0001$; odds ratio = 1.55. Actional: Log-likelihood = 13.13; $p < 0.001$; odds ratio = 1.29. Indeterminate: Log-likelihood = 89.45; $p < 0.0001$; odds ratio = 32.59.

Fig. 5 provides more information about the breakdown of the main categories into those subcategories as listed in Table 1.

As shown in Fig. 5, SAY-expressions are predominantly used with pragmatic functions (69%): organizing discourse, expressing opinions, or acting as interactional devices. This finding is consistent with the general trend that Goossens (1982) observes, namely that the communication verb SAY often takes on various pragmatic functions, ranging from textual to interpersonal, while the actional

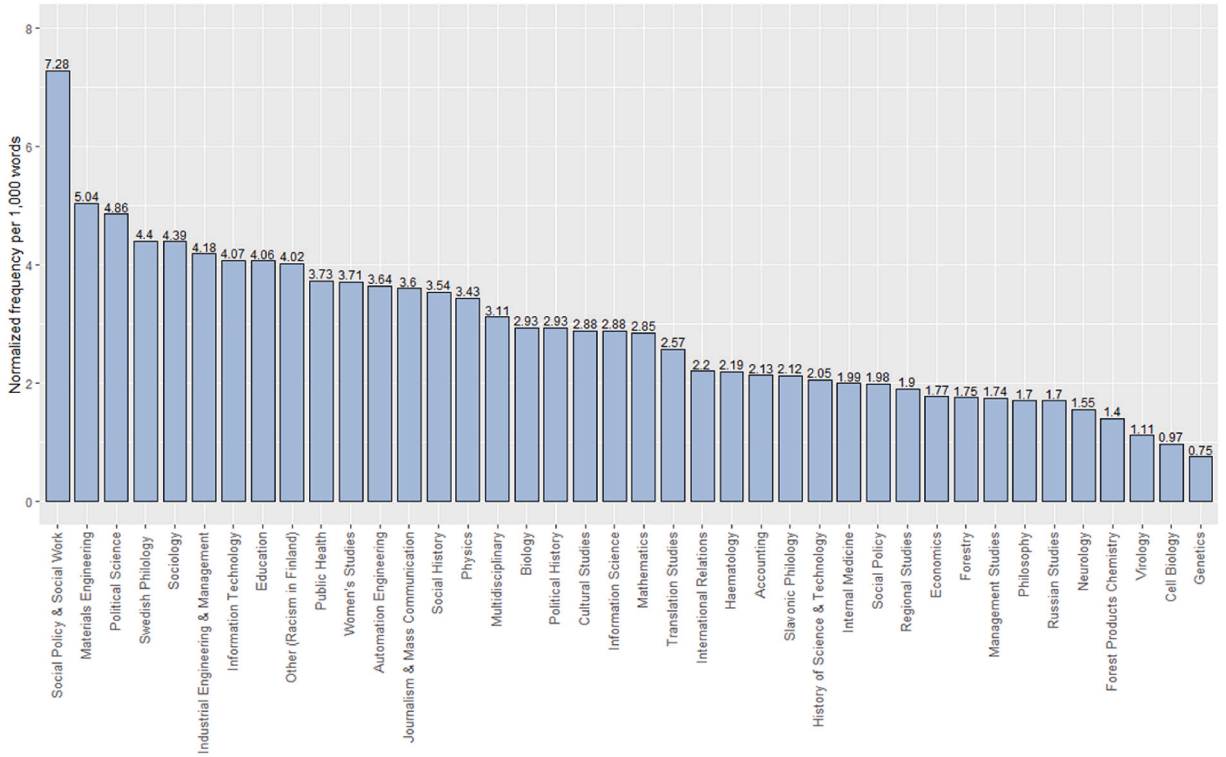

Fig. 3. Distribution of SAY-expressions across disciplines.

Table 2
Distribution of different forms of SAY.

	tokens	%	types	%
say	2226	63	1004	59
said	830	23	389	23
says	232	7	110	7
saying	251	7	188	11
TOTAL	3539	100	1691	100

Table 3
Most common SAY-expressions and their raw frequencies (≥10 for *say, said, says*; ≥5 for *saying*).

say	*said*	*says*	*saying*
let's say (361)	*as I said* (71)	*he says (that)* (43)	*sth saying (that)* (13)
I would say (that) (200)	*he said (that)* (61)	*it says (that)* (19)	*you're saying (that)* (7)
you say (that) (120)	*as you said* (30)	*X says (that)* (23)	*I'm saying (that)* (9)
they say (that) (69)	*X said (that)* (35)	*it says here (that)* (10)	*I'm not saying (that)* (7)
so to say (40)	*they said (that)* (48)		*they're saying (that)* (7)
I have to say (that) (24)	*you said (that)* (49)		*what you're saying* (5)
you can say (that) (17)	*I said (that)* (31)		*he was saying (that)* (5)
we can say (that) (35)	*as I said before/earlier/at the beginning* (17)		
I say (that) (21)	*as X said* (15)		
as you say (15)	*like I said* (14)		
when you say (that) (14)			
that is to say (that) (14)			
you might say (that) (13)			
I must say (that) (13)			
you could say (11)			
we could say (that) (10)			

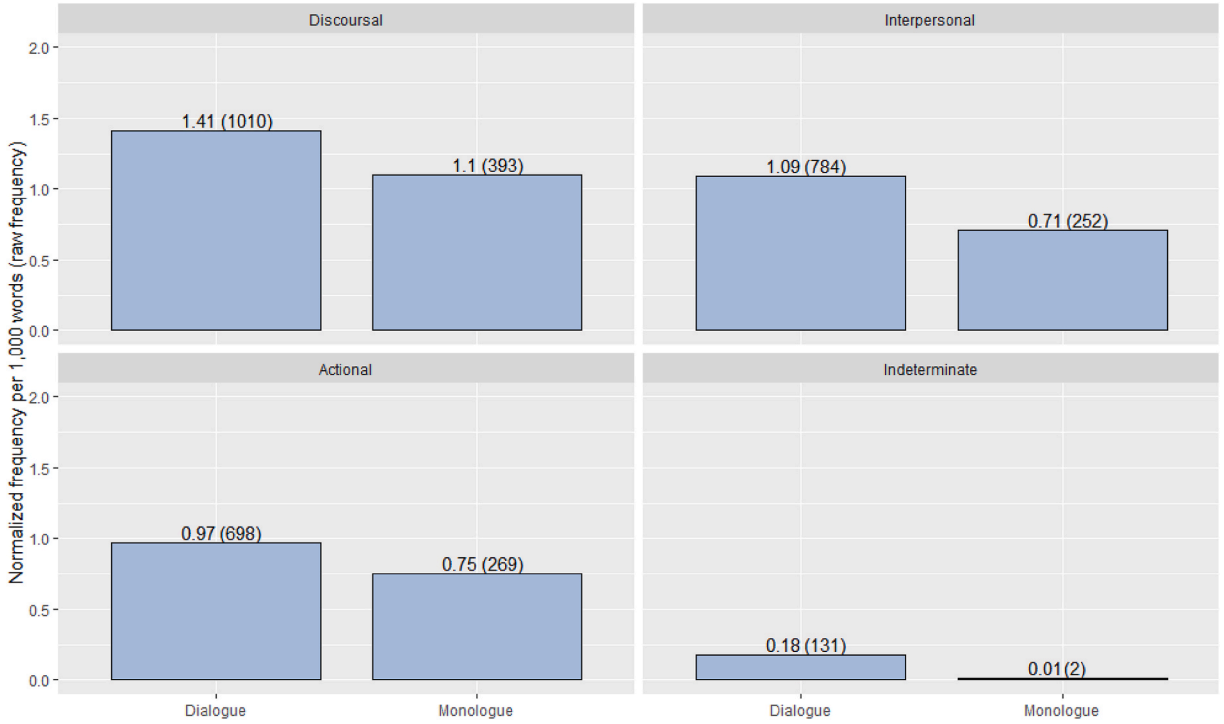

Fig. 4. Distribution of functions across dialogic and monologic event types.

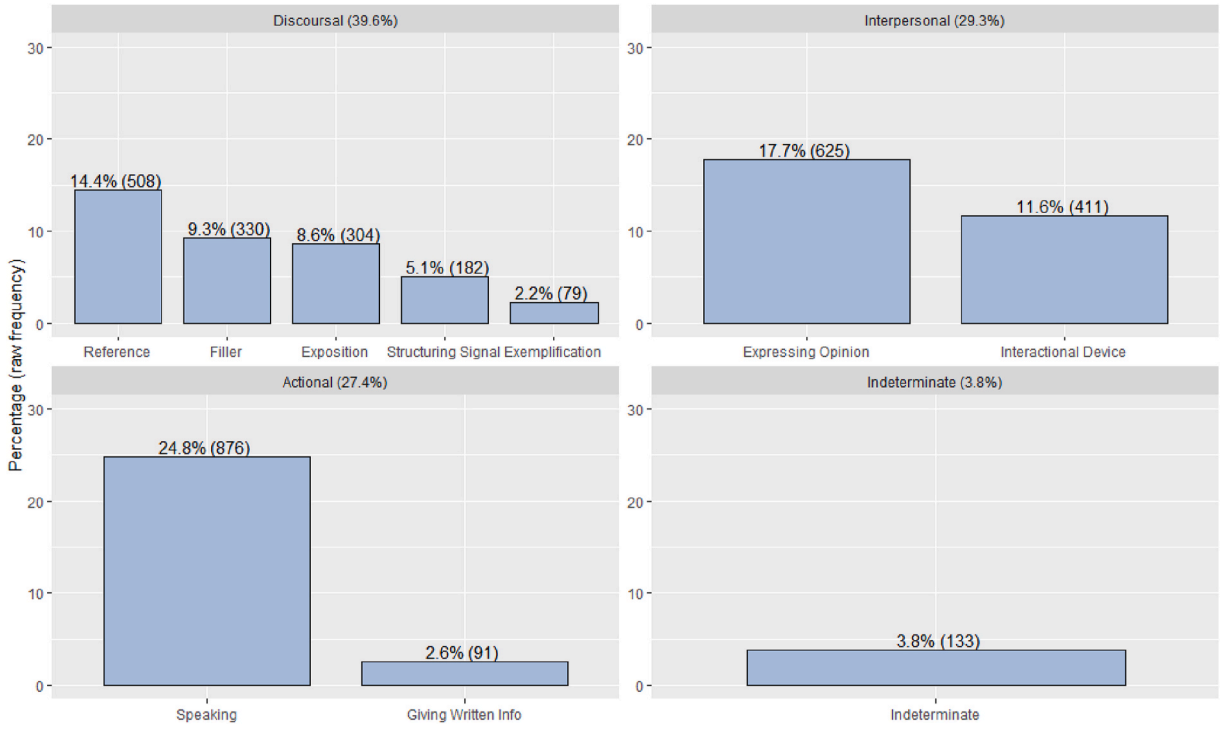

Fig. 5. Distribution of functions (broken down into subcategories).

meaning of the verb is largely backgrounded or marginalized.

Brinton (2008) provides an overview of the pragmatic functions of *say* in present-day English, focusing on a number of constructions/categories: *I say, let's say, I dare say,* (*as*) *you say*, and *that is to say*. In what follows, the main functions emerging from the present study are presented in comparison with Brinton's categories.

Among the pragmatic uses of SAY-expressions found in the ELFA corpus, the most common one may be regarded as belonging to the category of (*as*) *you say* as discussed in Brinton (2008), used to remind the hearer(s) of what someone has said on a previous occasion to either take issue with the information or to confirm understanding or interpretation. A number of formulaic sequences are associated with this function, particularly *as I said* and *you say* (*that*). Each has various forms; for the former, the conjunction slot can be filled with either *as* or *like*, and the subject slot *I* can be replaced by other pronouns or personal names e.g., *as X said* (*that*), *as you said*, *like I said*, *like you said*; occasionally the verb can also occur in the present tense e.g., *as you say, like you say*. Intervening or additional elements, typically in the form of time or place adverbials, were also found, e.g. *as sb said here/in the last session, as I said earlier/here/at the beginning*. The high frequency of this function is clearly attributable to the academic settings in question. Some expressions are typically used in academic lectures and presentations (e.g., *as I said earlier/here/at the beginning*), and some in discussions (e.g., *like/as you said*).

The expression *I dare say* was not found in the present study. However, its function, i.e., expressing speaker tentativeness of the truth value of the matrix clause, typically glossed as 'I suppose, assume, presume', turned out to be the second most common function of the pragmatic uses of SAY-expressions in the ELFA corpus. Brinton (2008) mentions some variant forms of *I dare say* such as *I must say, you might say*, which were also found in the present study. But the more frequently used forms are *I would say* (*that*), followed by *we can* (*also/of course/really*) *say* (*that*). The other, less common formal variants include *we could say* (*that*), *you could say* (*that*), *you might say* (*that*), *I will say* (*that*), *one could* (*also*) say (*that*).

Let's say is the most frequent SAY-expression in the data, but its main function, which has the sense of 'for example', according to dictionaries and the literature (e.g., Brinton, 2008; Van Olmen, 2013), accounts for about one fourth of the occurrences. Multiple functions were found in the remaining instances (cf. Section 4.3).

I say occurs sporadically in the data, and was not used with its main pragmatic function as discussed in Brinton (2008), namely attracting attention or expressing surprise. It was used in the ELFA corpus with its primary meaning, often followed by direct speech, arising from the need of narrating an event, as in the following example.

(7) *and then recently i met him and* ***i say*** *so now what do you think and* ***he says*** *listen he says i i've not changed my mind radically* (CPRE06C)

Other expressions which have the same function but occur more frequently include *he said* (*that*), *he says* (*that*), *they said* (*that*), *they say*, *X said* (*that*), *X says* (*that*), *I said*, with *he* and *they* dominating the pronoun slot, and the tense choice varying between the simple present and the simple past. In fact, this is the most common function associated with SAY-expressions in the present study. Again, this may be a feature of academic discourse in general, of which citation practices constitute an essential part. The frequency of this function is particularly high in one specific event type, namely thesis defense discussion, where there is a need to constantly refer to what is said/written in the thesis under discussion.

That is to say, which is a textual device often employed in written academic discourse, occurs only a handful of times in the spoken data under investigation. The function of elaborating on what has just been said is more often fulfilled by *let's say*. Another set of expressions involve the pattern *I* + progressive SAY: *I'm* (*not*) *saying* (*that*), *what I'm saying* (*here*) *is* (*that*).

Functions that are not covered in Brinton (2008) but cannot be ignored in the ELFA corpus belong to two main categories: filler and interactional devices. As regards the former, again, it is *let's say* that is the dominant form. Another fixed expression *so to say* was also often used as a filler. In addition, there are various forms of the pattern *how* + *say*: *how to say* (*it/that*), *how should I say* (*it/this*), *how do I say, how could/can I say, how do you say, how you say* (cf. Section 4.3).

With respect to interactional devices, question forms involving SAY are often used to elicit response from the addressees, e.g., *what do you say*, *what would you say*, *would you say that*, *can you say that*, *can you say something*, *did you say that*. Such questions can take various forms, most of which occur only once or twice; in other words, no single expression seems to be preferred by the speakers as a whole. However, when it comes to phrases used to respond to questions or other comments, there seem to be clearer preferences. In addition to *I would say* and *let's say*, which were commonly used to begin an answer, the other phrases were employed either when the speaker did not know the answer, e.g., *I can't* (*really*) *say* (*exactly/much*), *it's hard/difficult* (*for me*) *to* say, or when the question addressed implies some kind of criticism, e.g., *I have to say* (to acknowledge the criticism) (see (5)).

4.3. Discussion

From the results presented above it is clear that apart from *let's say* and *so to say*, idiomatic expressions in the target language as studied in previous studies and listed in dictionaries, such as *not to say, I'll say* (to express agreement), *you can say that again*, *I dare say, needless to say* rarely occur in the present study. The use of established formulaic sequences in the target language, particularly those that involve metaphorical meanings or pragmatic functions that cannot be directly derived on the basis of their components, requires a common linguistic and cultural background among speakers. In lingua franca communication where there is no such common ground, it is thus understandable that speakers would rely primarily on regular grammar either due to their limited access to prefabricated units in the target language or as a result of a conscious effort to stay away from idiomatic units that can cause difficulty for interactants without a common L1 background. Indeed, previous studies have observed approximation of formulaic language in ELF communication, featuring often the literal meaning over the more idiomatic but less transparent one (e.g., *I'll call you* ***later*** instead of *I'll call you*

back) (e.g., Bardovi-Halig, 2009). The types of SAY-expressions employed by the ELF speakers in the present study seem therefore consistent with what has been observed in the literature.

On the surface, the various formal configurations suggest that the open-choice principle is in operation, generated anew at the time of use. However, looking more closely at them together, interesting features of formulaicity can be found. In what follows, some of these features are summarized and discussed with corpus examples.

4.3.1. Different forms, same function

As presented in Section 4.2, each function was realized by a range of SAY-expressions, which, in turn, can have a number of variant forms. Out of all the main functions, the filler function is not covered in the literature dealing with the verb SAY and is therefore most likely to be typical of ELF communication. Both idiomatic and what seem to be freely generated expressions were frequently used as fillers for the speakers to facilitate their thought process as in the following examples:

(8) *so there are like many er er a whole set er a complex**, i don't know how to say** (it) er a set of er reasons …* (ULECD030)
(9) *so it's **how i would say** so there's a like er reaction times rea-reaction time measures with continuous and then it is er translated to the logarithm which describes this* (ULECD110)
(10) *they are not er they are not you don't learn any er the er or how **how should i say** they are not very, pedagogic or whatever* (USEMD03A)
(11) *… that's what we mean by public spiritedness and also a kind of, **you might say** a willingness to erm to involve yourself in these public issues without regard for your own self-interest* (ULEC210)
(12) *… the post-structuralist discourse is somehow the continuing the **so to say** natural continuation of a secularized western discourse* (UDEFD110)
(13) *and now it's certainly in a number of countries certainly in holland probably also in denmark er the media don't feel inhibited anymore to erm to er associate or yeah **let's say** associate immigrants with er with er things like crime suppression of women intolerance* (ULEC210)

In some cases (e.g. *how do I say, I don't know how to say, how I would/should say*), the filler function is fairly clear from the literal meaning of the expressions themselves, giving support for the view that L2 users rely on regular grammar to create formulas of their own (Kecskes, 2019). These expressions do not add anything to the proposition in question, but are there simply to help the speaker gain time to think of the right word(s), as in (8), or to reformulate an idea that was somehow abandoned midway, as in (9) and (10). The first personal pronoun can be replaced by *you* as in (11). The same function was also found in the use of the more idiomatic expressions such as *so to say* and *let's say* as in (12) and (13). Many of the filler expressions occur in the middle of a noun phrase, ahead of the head noun, and the speaker's thought process becomes even more evident given the accompanying repetitions and repairs in different configurations.

4.3.2. Same form, different functions

While different forms can be used with the same function, the same form can also be used with different functions. Take *I would say* as an example. It was often used in association with its prototypical function, namely to give an opinion with a certain degree of uncertainty and willingness to be corrected, as in (14). Apart from this function, it was also used as a kind of interactive device to express agreement, as in (15), or to elicit an answer in response to a question raised in the preceding discourse, as in (16). It occurs twice in (17). The first instance is similar to that in (16), while the second one is somewhat different as no further question was asked, but the speaker used the expression to start the next turn in conversation. The example in (18) illustrates the use of *I would say* to further clarify what has been said.

(14) *… this is a challenge i **would say** and i think he succeeds …* (UDEFD140)
(15) <S1> *… [do you] mean to cover relations* </S1>
<S2> *yeah yes **i would say** i* <S1> *[(xx)]* </S1> *[(xx)] yes yeah okay yes …* (USEMD120)
(16) *<S3> … so what erm what then can other regions learn from your from from your findings </S3>*
<S2> *erm **i would say** er first of all networking methods …* (UDEFD020)
(17) <S3> *mhm, if you split the er the er work load between the two what would you say the split would be* </S3>
<S1> ***i would say that** maybe that er the initial design and the whole let's say design process would be maybe 25 per cent of all and the rest would be implementation* </S1>
<S3> *mhm mhm good that's good to know* </S3>
*<S1> but **i would say that** there were also quite many ideas that came up in the implementation …* (UDEFD080)
(18) *so it might be interesting to get some comments how people see if they have been able to create some contacts which they think would be useful want to invite others to work on some specific er topics or items and also any feedback on the seminar days we've had now er how we could improve these kind of er gatherings what could we do for the future. so i **would say** it's sort of open floor for comments and suggestions.* (CDIS08B)

The examples above were produced by different speakers. It is thus possible that the expression has been picked up as a whole entity from the speakers' experience with the language. The fact that one speaker tends to associate the expression with a specific function that differs from the usage by another speaker is in line with Vetchinnikova's (2015) observation of the 'recycling' of ready-made word sequences in ELF communication. The ready-made sequences are recycled in the sense that the surface structure is preserved intact, but

the use is marked by a measure of creativity in terms of their communicative function to suit the speakers' own needs. More examples of the same expression being used for different functions can be found in the following discussion in relation to personal preference.

4.3.3. Personal preference

It emerged from our coding that many speakers have a clear preference for a set of individual expressions. The following examples, all produced by the same speaker, show that the expression *let's say*, which is typically associated with the sense 'for example' in English (Brinton, 2008), was used for a variety of purposes.

(19) *it's when you are measuring adhesion is always a problem for any for any type of systems because you are measuring very small values* ***let's say*** *you are working with things like 0.4 up to 1.4 MPa,* …(UDEFD010)
(20) *it's very applied* ***let's say*** *there is a very clear industrial aim in the thesis* … (UDEFD010)
(21) … *sometimes the scientific level is not er so good as the as the applied level* ***let's say,*** …(UDEFD010)
(22) *then you have the resins that are like* ***let's say*** *the the muscles in that will provide the strength to our to our body* … (UDEFD010)
(23) *if you go to page nine* ***let's say*** *the adhesive films were prepared by melting the polymer for 15 min at 160 degrees C* (UDEFD010)

The example in (19) is a clear case where the conventional function of *let's say* is involved, clarifying what *very small values* means by giving concrete examples. In (20), *let's say* introduces an explanation of what *applied* means. It was also found in connection with clearly evaluative terms, such as *not so good* in (21), to express personal opinions. In (22), *let's say* seems redundant – the sense of 'for example' is already delivered by the word *like*; the repetition (*the the*) that follows *let's say* gives a further indication that the speaker was looking for the right word (*muscles*) while using filler words and phrases to fill the void. In (23), *let's say* was used to introduce a quotation from a doctoral thesis under discussion. In a way, it may be regarded as having the same function as *you say*, but the speaker here displayed a particular preference for the expression *let's say*. Indeed, the multiple functions of the expression seem to suggest that it may have become what Hasselgren (1994) calls a phrasal 'teddy bear' that the speaker clings on to regardless of circumstances.

Different L2 speakers may have different teddy bears. The following examples show how *we might say* was used by another speaker with multiple functions: filler in (24) and attitude marker in (25). However, unlike *let's say*, which is invariable in form, the teddy bear in this case can take slightly different forms where the subject slot can be filled by other pronouns such as *you* and *one*. The expression *you might say* in (26), for instance, has exactly the same function as *we might say* in (25): both occur together with an evaluative term (*opportunity*, *difficult*) and have the same sentential position (at the end of the proposition).

(24) … *the im-the immigrant culture still serves functions such as erm er* ***we might say*** *providing a kind of psychological shelter for them* … (ULEC210)
(25) *it gives immigrants job opportunities gives them erm an opportunity to er to be-become upwardly socially mobile* ***we might say*** (ULEC210)
(26) *it's er always quite different difficult to distinguish precisely between these various strands* ***you might say*** … (ULEC210)

The following examples were produced by a different speaker in a presentation. In addition to some formal variability (*what* vs. *as*), personal preference is also evident in terms of function (textual reference), tense/aspect (present perfect), and sentential position (mid-VP).

(27) *we have also many nations with different size of economies cultures and this has had er positive st-er that's* ***what i have said*** *a positive strategy to regional integration that* … (CPRE05I)
(28) *erm latin american countries have not had reached out* ***as i have said*** *to regional integration due to this* … (CPRE05I)
(29) *we have two developed countries there and erm the rest we consider developing countries there is erm this effort to unite the economies* ***as i have already said*** *to form this free trade area er* … (CPRE05I)

The same can be said of the SAY-expressions in (30) and (31), with the use of the progressive aspect and the accompanying discourse marker *like*. All the expressions were used to introduce direct or indirect speech, attributing a proposition to someone else.

(30) … *i remember like at one time er* ***president was saying*** *like that er the country has er approximate of 36 million people and* ***he was saying*** *like almost if the debt is distributed to each and every person in the country* … (USEMD100)
(31) *HIV is like because it's it's global thing and* ***everybody's just saying*** *like yeah HIV's killed a lot of people* … (USEMD100)

To sum up, personal preference is not only associated with invariable expressions; it can also manifest itself through formal variations. In the latter case, taken individually, the various formal configurations may appear to be freely generated according to the open-choice principle. However, when they are put together, the same underlying pattern (in terms of tense/aspect choice, sentential position, accompanying element(s), function) suggests that the pattern itself may be treated as a holistic entity where there are both fixed elements and open slots to be filled freely.

5. Conclusion

Focusing on one word has made it possible to provide a comprehensive picture of its phraseological patterns in the one-million-

word ELFA corpus, thereby shedding further light on features of formulaicity that characterize academic ELF communication and their relationship with a range of factors such as event type, discipline, and individual preference.

To recap the main findings, first of all, non-compositional expressions in the target language rarely occur in the present investigation, corroborating a well-observed tendency among ELF speakers in general to use transparent expressions (Bardovi-Halig, 2009; Kecskes, 2019; Pitzl, 2012). The fact that most SAY-expressions produced by ELF speakers in the present study can be broken up into their component parts may suggest the operation of analytical processing, which L2 speakers are said to depend on primarily (Pawley & Syder, 1983; Sinclair, 1991; Wray 2008, 2012). However, taken together, it could also be argued that those seemingly transparent expressions display formulaic features that point to a holistic approach to a certain extent. To begin with, some expressions that are commonly seen in the target language (e.g., *let's say, I would say*) are often used verbatim, but with variations in function. These expressions may be, to use Vetchinnikova's (2015) term, 'recycled' directly from L2 users' experience with the target language in terms of their surface structure, but they are used with a certain degree of 'creativity' in terms of their communicative function to meet the speakers' own needs at the moment of use, such as the need to gain time without relying too much on traditional filler words (e.g., *like*, *you know*), which may diminish their communication style. In other words, it is also possible that L2 users have memorized these sequences that they are exposed to as whole entities, but use them in their own ways. Indeed, throughout the study, there is evidence suggesting not only that the same expression can be used with different functions, even by the same speaker, but the same function can also be fulfilled by a wide variety of expressions.

Even with those expressions that display formal variability, it is questionable whether they can always be taken as novel combinations built from individual words and grammatical rules. Amid various formal configurations produced by individual speakers (e.g., *we/you might say, he/everybody/president was saying like*), certain elements remain stable (such as tense choice, underlying syntactic structure). More importantly, each underlying syntactic pattern seems to be tied up with a fixed function (e.g., *X might say* to express opinion, *X was saying like* to introduce a direct or indirect quotation). All these features seem to be in line with the theories about formulaic language use in ELF communication such as approximation and fixing (Mauranen, 2012; Vetchinnikova, 2015), both of which suggest holistic language processing. While approximation or fixing often involves some local loosening of a non-compositional expression (e.g., *carved in stones* vs. *carved in stone*) (Pitzl, 2012), the present study suggests that the holistic approach is also likely to be in place with those that appear to be rule-governed expressions, which for some reason may have become fixed to a certain extent among individual speakers. The 'loosening' in turn may be seen as taking the form of a few open slots in the pattern, which may be filled according to the open-choice principle. The use of SAY-expressions therefore suggests that phraseological patterning in ELF may well be a result of both holistic and analytical mechanisms working together.

Overall, these features seem compatible with what is reported in the literature about ELF communication. However, without a comparable non-ELF corpus, which is a major limitation of the present study, it remains to be validated with empirical evidence whether these features are indeed unique to ELF. For that purpose, the Michigan Corpus of Academic Spoken English (Simpson et al., 1999), which contains data produced by both L1-and L2-English speakers in a variety of event types and disciplines, would be a potentially useful comparison corpus for future research. Notwithstanding the need for further investigation, the versatility of SAY-expressions as demonstrated by the ELF speakers under study certainly warrants more attention to those formulaic expressions that are perhaps less salient than what are commonly taught in EAP classes and their role in facilitating spoken academic communication in the teaching of EAP.

On balance, it seems that event types and individual preferences are more likely to affect the frequency and use of SAY-expressions than academic disciplines. This does not mean that discipline is an irrelevant variable to investigate in research on spoken academic discourse, but it does raise the question if it is feasible to look at the effect of one variable independent of the others. In addition, as the present study involves one communication verb SAY and its phraseological patterns, the nature of the verb may decide that event types and even individual preferences are more likely to influence their frequency and usage. This makes it even more complicated if we are to establish the relationship between formulaic language use in general, including formulaic sequences that involve different key words, and relevant contextual and individual variables. In that sense, the present study has given rise to further questions for future research to explore.

Author statement

The first author (Y. Wang) designed and conducted the study as well as drafted, revised and edited the manuscript. The second author (H. Kaatari) created all the figures, conducted statistical tests and reported the results, and revised and edited the manuscript.

References

Ädel, A. (2008). *What uh the folks who did this survey found*: Expert attribution in spoken academic lectures. *Nordic Journal of English Studies, 7*(3), 83–102.
Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81–92.
Anthony, L. (2020). *AntConc* (Version 3.5.9) [Computer Software]. Tokyo: Waseda University. Available from https://www.laurenceanthony.net/software.
Bardovi-Halig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning, 59*, 755–795.
Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.
Brinton, L. J. (2008). *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge: CUP.
Buerki, A. (2016). Formulaic sequences: A drop in the ocean of constructions or something more significant? *European Journal of English Studies, 20*(1), 15–34.
Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397–423.
Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165–193.

Gablasova, D., Brezina, V., McEnery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics, 38*(5), 613–637.
Goosens, L. (1982). *Say*: Focus on message. In R. Dirven, L. Goosens, Y. Putseys, & E. Vorlat (Eds.), *Pragmatics & beyond, 3:6The Scene of Linguistic Action and its Perspectivization by speak, talk, say and tell, 85-131*. Amsterdam: John Benjamins.
Halliday, M. A. K., & revised by C.M.I.M. Matthiessen.. (2014). *Halliday's introduction to functional grammar* (4th ed.). Oxen: Routledge.
Hesselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237–258.
House, J. (2009). Subjectivity in English as Lingua Franca discourse: The case of *you know*. *Intercultural Pragmatics, 6*(2), 171–193.
Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4–21.
Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41–62.
Kashiha, H., & Heng, C. S. (2014). Discourse functions of formulaic sequences in academic speech across two disciplines. *Journal of Language Studies, 14*(2), 15–27.
Kecskes, I. (2019). Formulaic language and its place in intercultural pragmatics. In A. Siyanova-Chanturia, & A. Pelicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective*. London: Routledge.
Mauranen, A. (2008). *The Transcriptions of the ELFA corpus, downloadable version* [text corpus]. *Kielipankki – The Language Bank of Finland*. http://urn.fi/urn:nbn:fi:lb-2014052721.
Mauranen, A. (2009). Chunking in ELF: Expressions for managing interaction. *Intercultural Pragmatics, 6*(2), 217–233.
Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: CUP.
Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes, 29*(3), 183–190.
Metsä-Ketelä, M. (2012). Frequencies of vague expressions in English as an academic lingua franca. *Journal of English As a Lingua Franca, 1*(2), 263–285.
Pan, F., Reppen, R., & Biber, D. (2020). Methodological issues in contrastive lexical bundle research. *International Journal of Corpus Linguistics, 25*(2), 215–229.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards, & R. Schmidt (Eds.), *Language and communication, 191–226*. London: Longman.
Pitzl, M.-L. (2012). Creativity meets convention: Idiom variation and re-metaphorization in ELF. *Journal of English As a Lingua Franca, 1*(1), 27–55.
Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction, 3*(1), 95–119.
Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: OUP.
Shinzato, R. (2004). Some observations concerning mental verbs and speech act verbs. *Journal of Pragmatics, 36*, 861–882.
Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (1999). *The Michigan corpus of academic spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: OUP.
Swales, J. M. (2019). The future of EAP genre studies: A personal viewpoint. *Journal of English for Academic Purposes, 38*, 75–82.
Van Olmen, D. (2013). The imperative of *say* as a pragmatic marker in English and Dutch. *Journal of German Linguistics, 25*(3), 247–287.
Vetchinnikova, S. (2015). Usage-based recycling or creative exploitation of the shared code? The case of phraseological patterning. *Journal of English as a Lingua Franca, 4*(2), 223–252.
Wang, Y. (2017). Lexical bundles in spoken academic ELF: Genre and disciplinary variation. *International Journal of Corpus Linguistics, 22*(2), 187–211.
Wang, Y. (2018a). Formulaic sequences signalling discourse organisation in ELF university lectures: A disciplinary perspective. *Journal of English as a Lingua Franca, 7*(2), 355–376.
Wang, Y. (2018b). *As Hill seems to suggest*: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *Journal of English for Academic Purposes, 33*, 12–23.
Wang, Y. (2020). Methodological challenges in identifying formulaicity in spoken academic ELF communication. In Kumiko Murata (Ed.), *ELF Research and Approaches to Data and Analyses: Theoretical and methodological Concerns* (pp. 143–157). Oxon: Routledge.
Wray, A. (2002). *formulaic language and the lexicon*. Cambridge: CUP.
Wray, A. (2008). *formulaic language: Pushing the boundaries*. Oxford: OUP.
Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics, 32*, 231–254.

Ying Wang is associate professor in English linguistics at Karlstad University in Sweden, where she teaches academic writing and linguistics courses. Her main research interests lie in areas of corpus linguistics and EAP, with a focus on formulaic language use in spoken and written academic discourse.

Henrik Kaatari received his PhD from Uppsala University in Sweden. His research interests include syntactic/grammatical variation, complexity and learner corpus research. He is currently senior lecturer in English linguistics at the University of Gävle in Sweden.