Criterial Grounded Canonical Constructions in an L2 Multimodal Spoken Interaction Learner Corpus

Yoshiyuki Notohara
Doshisha University
ynotohar@mail.doshisha.ac.jp

Abstract

This study examines whether grounded canonical constructions are criterial in an L2 multimodal spoken interaction learner corpus. While corpus pragmatics research has explored the conventionalised illocutionary force salience of markers and formulaic sequences (Rühlemann, 2019), sentence-level construction salience remains empirically unexplored (Ruytenbeek et al., 2017). Cognitive usage-based linguistics frames construction-illocutionary force relationships as *metonymic inferencing* from illocutionary scenarios (Panther, 2022; Stefanowitsch, 2003); the Location schema construction with modality (e.g., "I'll be there tomorrow") may function as a promise [STATING FUTURE PLANS FOR PROMISING]. Such grounded, tense-aspect-modality (TAM)-specified canonical constructions are conventionally entrenched in repeated spoken interaction (Schmid, 2020). The current study investigates whether two grounded, TAM-specified Material World canonical event schema constructions, which are infrequent and semantically peripheral in native data, are criterial across CEFR levels: (a) the Processes schema construction (e.g., *I'm getting better*) and (b) the Object-motion schema construction (e.g., *The prize goes to a child*). With only limited accuracy, the Random Forest (RF) and the Bayesian Additive Regression Trees (BART) analyses identified the Processes schema constructions as candidate criterial features: (a) present progressive with comparatives [DESCRIBING A GRADUAL ONGOING CHANGE FOR EXPRESSING THE SPEAKER'S SUBJECTIVE ASSESSMENT] (B1_2); (b) with past tense and negative adjectives [DESCRIBING A PAST CHANGE OF STATE FOR REPORTING A PAST NEGATIVE MENTAL STATE] (B1_2); (c) with the present tense and negative adjectives [DESCRIBING A PRESENT CHANGE OF STATE FOR REPORTING A PRESENT NEGATIVE MENTAL STATE] (B2+). No critical findings emerged for the Object-motion schema construction.

Keywords: canonical constructions, corpus pragmatics, spoken interaction learner corpus

1. Introduction

Since the 1960s, researchers in spoken English or spoken discourse analysis have studied lexis and lexico-grammatical patterns using small spoken corpora. From the late 1990s to the early 2010s, with the development of a larger historical and contemporary spoken corpora, researchers in modern English grammar and pragmatics studied and referred to pragmatic functions of word classes (e.g., verbs, adverbs, and auxiliaries), grammatical categories (e.g., tense, aspect, and modality), discourse markers (e.g., *actually*), conversational routines (e.g., ADVERB *sorry to*), and speech acts (e.g., *Why don't you* X?) (Adolphs, 2008; Aijmer, 1996, 2013; Biber et al., 1999, 2021; Carter & McCarthy, 2006; Huddleston & Pullum, 2002; Jucker & Taavitsainen, 2013). Quite recently, in the late 2010s, researchers in corpus pragmatics have investigated the conventionalised illocutionary force salience of markers and formulaic sequences using audio-based and multimodal spoken corpora mostly under *the form-to-function approach* (Aijmer & Rühlemann, 2015; Rühlemann, 2019), even though there are a few *function-to-form approach* with spoken corpora with pragmatic annotations, such as the DART taxonomy (Weisser, 2015; 2020).

Since the early 2000s, philosophies of language researchers have focused on the pragmatic functions of syntactic constructions, specifically, sentence-level constructions, while critically engaging with classical Searlian pragmatics (Alston, 2000). Researchers in *cognitive pragmatics*, influenced by post-Gricean pragmatics (Sperber & Wilson, 1986, 1995; Mazzone, 2019), have revised classical Searlian pragmatics, analysing construction-illocutionary force relationships as *pragmatic (metonymic) inferencing* from *illocutionary scenarios* under *felicity conditions* (Panther, 2022; Panther & Thornburg, 1998, 2003, 2017; Thornburg & Panther, 1997; Stefanowitsch, 2003). For example, the Location schema construction with modality, as in "I'll be there tomorrow," may function metonymically as a promise [STATING FUTURE PLANS FOR PROMISING]. Such *grounded, tense-aspect-modality (TAM)-specified canonical constructions* are assumed to be "pragmatically grounded and developed through social interactions" (Langacker, 2008, 2009). Furthermore, the constructions are conventionally entrenched in repeated spoken interaction (Schmid, 2020). In fact, few empirical studies in cognitive pragmatics have examined this *syntax-semantics-pragmatics interface*, namely, construction-illocutionary force relationships in corpus pragmatics, although theoretical frameworks have so far been discussed in cognitive pragmatics (e.g., Panther & Thornburg, 2003; Panther, 2022).

Based on this theoretical background, this paper first of all briefly reviews a previous empirical study on construction-illocutionary force relationships in three of the five TAM-specified Material World constructions; for example, the State schema construction with the present tense (e.g., *This is true*); the Location schema construction with the present tense (e.g., *I'm here*); the Possession schema construction with the past tense (e.g., *I had a brilliant idea*) (Notohara, 2025). Second, it reports novel results on two unexplored grounded canonical constructions: (a) the Processes schema construction (e.g., *I'm getting better*) and (b) the Object-motion schema construction (e.g., *The prize goes to a child*) through the Spoken BNC2014 corpus. Third, it also examines whether the two TAM-specified Material World canonical constructions, which are infrequent and semantically peripheral in native data, are criterial across CEFR levels using the L2 learners' multimodal spoken dialogue (SD) component of ICNALE (Ishikawa, 2023).

2. Literature Review

2.1 Canonical Event Schema Constructions

Radden and Dirven (2007) analysed sentence patterns in terms of thematic roles (or case roles, semantic roles) and summarised the typical configurations of thematic roles into 11 canonical event schema constructions (CESCs: constructions 1-7, 9-12 in Table 1), grouped under Material World, Psychological World, and Force-Dynamic World. Through the spoken component of the BNC corpus (1994) and the FrameNet (Fillmore, Johnson & Petruck, 2003), Notohara (2014) examined *the canonicity* of the 11 CESCs (i.e., whether each construction is frequent and closely related to the superordinate schema of the inheritance relationships between constructions) and extended 11 to 13 CESCs by adding the mental schema construction (Psychological World) and the Communication schema construction (Force-Dynamic World) (CESCs: construc tions 8 and 13 in Table 1).

Later, Notohara (2016) confirmed *the Zipfian type-token verb frequency distribution* within 13 CESCs and confirmed 13 frequent verbs for the extended 13 CESCs through the grammatically annotated adult native English corpus, the ICE-GB R2 Corpus (2006). Notohara (2021) replaced *want* with *like* as the prototypical verb of the Emotion schema construction because *want* exhibits more skewed construction patterns with *to*-infinitive than *like*, modifying two examples to make them clearer in the Possession schema (CESCs: constructions 5 and 6 in Table 1).

CESCs are commonly used even in a daily conversation such as small talk on popular TV entertainment as follows:

01	Jo:→	***I once broke my ankle***, (.)	9. Action
02		but instead of calling an ambulance,	
03		→ ***I went all the way to A&E***	10. Self-motion
04		(0.6) *in a wheelbarrow*.	
05	Rob:	David's team? ↑	
06	David:	Oh. =	
07		= so, where were you ↑ (.)	
		when you broke your ankle? ↑	
08	Jo: →	***I was in the park***.	3. Location

("*Would I lie to you?*" Series 16:11 More Unseen Bits on BBC iPlayer 14 April 2023)

Figure 1: Canonical event schema constructions in talk

For example, in Figure 1, "I once broke my ankle*"* in utterance 01 can be categorised as the Action schema construction with the past tense (Reporting). "I went all the way to A & E" in utterance 03, which can be recognised as the Self-motion construction (Reporting). Finally, in utterance 08, "I was in the park," which can be interpreted as the Location schema construction with the past tense (Reporting).

Notohara (2025) examined the association strengths and directions between three CESCs relating to Material World (the Occurrence schema: States, the Spatial schema: Location, and the Possession schema) and frequent illocutionary forces in the Spoken BNC2014 corpus, using collostructional analysis, specifically, ΔP, one of several association measures available (See Dunn, 2022 for details). Similarly, Notohara (2023a, 2023b) examined canonical construction-illocutionary force relationships in Psycho-logical World and Force-Dynamic World constructions. Table 1 lists the main findings from the two previous studies (See Illocutionary forces in Table 1). Remaining unexplored canonical construction-illocutionary force relationships in CESCs relating to two Material World constructions are examined and presented in the following sections.

Event Schema	Participant Roles[a]	Sentence Patterns	Prototypical Verbs	Examples[b] (TAM)	Illocutionary Forces[c] (IFs)
Material World					
1.Occurrence schema: States	T-(T)	SVC	*be*	*This is true.*	Ide/Rep/Dis
2.Occurrence schema: Processes	T-(T)	SVC	*get*	*You'll get better.*	Rep/Pro/Cor
3.Spatial schema: Location	T-L	SV	*be*	*I'm here.*	Rep/Cor/Ask
4.Spatial schema: (Object)-motion	T-G	SV	*go*	*The prize goes to a child.*	Rep/Ask/Pro
5.Possession schema	P-T	SVO	*have*	*I have a book about her life.*	Rep/Cor/Ask
Psychological World					
6. Emotion schema	E-C	SVO	*like*	*He liked Hamburg and Münich.*	Rep/Cor/W&D
7. Perception/Cognition schema	E-T	SVO	*see*	*I saw a black cat on the step.*	Rep/Abi/Pro
8. Mental schema	E-T	SVO	*think*	*I think it is very true*	Cer/Cor/Ask
Force-Dynamic World					
9. Action schema: Energy chain	A-T	SVO	*make*	*I'll make some tea.*	Rep/Abi/Cor
10. Self-motion schema	A-G	SV	*go*	*This summer I went to Spain.*	Rep/Int/Pro
11. Caused-motion schema	A-T-G	SVO	*put*	*He put the platter on the floor.*	Rep/Int/Abi
12.Transfer schema	A-T-R	SVO	*give*	*She gave it to him.*	Rep/Abi/Obl
13. Communication schema	A-T-R	SVO	*say*	*He said it to me.*	Rep/AfC/Cor

Notes. [a]T = theme, L = location, G = goal, P = possessor, E = experiencer, C = cause, A = agent, R = recipient; [b]each example was extracted from the BNC Online (2003–2026); [c]Ide = identifying, Rep = reporting, Cor = correcting, Ask = asking, Pro = probability, W&D = wants & desires, Cer = certainty, Abi = ability, Int = intentions, Dis = dissatisfaction, Obl = obligations, AfC =ask for confirmation

Table 1: Extended 13 canonical event schema constructions (based on Radden & Dirven, 2007)

2.2 The Processes Schema Constructions and Conventionalised Illocutionary Forces

This section examines the construction-illocutionary force relationships in the ΔP scores between the TAM-specified Processes/SVC (GET) schema constructions (Material World), which remained underexamined, and illocutionary forces, as shown in Table 2.

L: CESCs	R: IFs	LR	RL	Freq
1.Processes/SVC(GET) +Past	Reporting	0.45	0.82	16
2.Processes/SVC(GET) +Progressive	Probability	0.31	0.52	8
3.Processes/SVC(GET) +Present	Probability	0.46	0.42	7
3.Processes/SVC(GET) +Progressive	Reporting	0.19	0.44	7
4.Processes/SVC(GET) +Perfective	Reporting	0.17	0.97	6
4.Processes/SVC(GET) +Modality	Probability	0.24	0.73	6
5.Processes/SVC(GET) +Present	Reporting	0.13	0.28	5

Notes. ΔP indicates the strengths and directions of illocutionary force salience from 0.00 to 1.00; LR shows left-to-right; RL shows right-to-left.

Table 2: Top 5 construction-illocutionary force relationship -s in the ΔP scores between the TAM-specified Processes/ SVC(GET) schema constructions and illocutionary forces.

Top three examples of strong construction-illocutionary force relationships identified in the data are in (1) to (3) as follows:

(1) Processes/SVC(GET) +Past→ Reporting
(Freq 16, LR: 0.45)
e.g., um so *I got very excited about this* and went home
(Spoken BNC2014)

(2) Processes/SVC(GET) +Progressive → Probability
(Freq 8, LR: 0.31)
e.g., yeah *it's getting boring* now shall we do that?
(Spoken BNC2014)

(3) Processes/SVC(GET) +Present → Probability
(Freq 7, LR: 0.46)
e.g., what are you gonna do when *you get older*?
(Spoken BNC2014)

In (1), *I got very excited about this*; a formulaic sequence often used in spoken conversation can be schematised as the Processes/SVC (GET) +Past [DESCRIBING A PAST CHANGE OF EMOTIONAL STATE FOR REPORTING A PAST EMOTIONAL EXPERIENCE]. In (2), *it's getting boring* can be identified as the Processes/SVC (GET) +Present Progressive [DESCRIBING ONGOING PROCESSES FOR EXPRESSING PRESENT PROBABILITY]. In (3), *when you get older* can be interpreted as the Processes/SVC(GET)+Present with comparatives [DESCRIBING A PRESENT GRADUAL CHANGE OF STATE FOR EXPRESSING PRESENT PROBABILITY].

2.3 The Object-motion Schema Constructions and Conventionalised Illocutionary Forces

Similarly, the construction-illocutionary force relationship -s in the ΔP scores between the TAM-specified Object-motion/SV (GO) schema constructions (Material World), which have not yet been examined, and illocutionary forces were examined and are shown in Table 3.

L: CESCs	R: IFs	LR	RL	Freq
1.Object-motion/SV(GO) +Past	Reporting	0.40	0.80	14
2. Object-motion/SV(GO) +Progressive	Asking	0.46	0.58	13
3.Object-motion/SV(GO) +Present	Reporting	0.25	0.45	9
3.Object-motion/SV(GO) +Modality	Probability	0.44	0.89	9
4.Object-motion/SV(GO) +Progressive	Reporting	0.21	0.34	8
5.Object-motion/SV(GO) +Present	Probability	0.24	0.25	5

Notes. ΔP indicates the strengths and directions of illocutionary force salience from 0.00 to 1.00; LR shows left-to-right; RL shows right-to-left.

Table 3: Top 5 construction-illocutionary force relationship -s in the ΔP scores between the TAM-specified Object-moti on/SV(GO) schema constructions and illocutionary forces

Top three examples from the data that illustrate strong con struction-illocutionary force relationships are shown in (4) to (6) below:

(4) Object-motion/SV(GO)+Past → Reporting
(Freq 14, LR: 0.40)
e.g., nearly at the end of it *my phone went*
(Spoken BNC2014)

(5) Object-motion/SV(GO)+Progressive → Asking
(Freq 13, LR: 0.46)
e.g., I'm not sure w*hat's going on?*
(Spoken BNC2014)

(6) Object-motion/SV(GO) +Modality → Probability
(Freq 9, LR: 0.44)
e.g., if you traced the chain *it would go everywhere* wouldn't it?
(Spoken BNC2014)

In (4), *my phone went* can be classified as the Object-motion/SV (GO)+Past [DESCRIBING A PAST CHANGE OF STATE FOR REPORTING A PAST UNEXPECTED DISRUPTION]. In (5), w*hat's going on?* can be analysed as the Object-motion/SV(GO)+Present Progressive [DESCRIBING AN ONGOING PROCESS FOR REQUESTING INFORMATION]. In (6), *it would go everywhere* can be construed as the Object-motion/SV (GO) +Modality with subjunctive mood [DESCRIBING A HYPOTHETICAL SPATIAL EVENT FOR DRAWING A LOGICAL INFERENCE].

2.4 Criterial Features in the English Profiles

Since the 2010s, researchers in the English Profile (EP) Project have studied a written learner corpus, the Cambridge Learner Corpus (CLC), and extracted lexicogrammatical patterns as criterial features (CFs) distinguishing adjacent CEFR levels from A2 to C2 (Hawkins & Buttery, 2010; Hawkins & Filipović, 2012). Littlemore, Krennmayr, Turner, & Turner (2014) examined metaphor density in L2 essay writing from CEFR A2 to C2 and extracted candidates for criterial metaphors: closed-class and open-class metaphors. Chen and Baker (2016)

investigated lexical bundles in a written learner corpus, the Longman Learner Corpus (LLC), and explored criterial discourse features from CEFR B1 to C1 in L2 learner expository and argumentative essays. Under a usage-based approach, Römer and Verger (2019) extracted L2 learners' English Verb-argument constructions (VACs) in writing texts from the Education First-Cambridge Open Language Database (EFCAMDAT). They examined the relationship between frequent verbs in L2 VACs and CEFR levels from A1 to C1, but unfortunately, they did not identify candidates for criterial features. Khusik and Huhta (2020) investigated syntactic complexity (e.g., subordination, coordination, phrasal sophistication, and phrasal density) in L2 argumentative essay writings from CEFR A1 to B1.

Regarding spoken corpus studies, Galaczi (2014) explored topic development organization, listener support moves, and turn-taking management in a learner corpus in L2 paired speaking tests and extracted criterial interactional features across CEFR levels from CEFR B1 to C2 through frequency analysis (FA) and conversation analysis (CA). Kang and Wang (2014) broadly examined criterial linguistic features (e.g., discourse management, grammatical resources, lexical resources, and pronunciation) and criterial interactive features (e.g., cooperation, coherence, turn-taking, and strategies) in L2 spoken interaction tasks across CEFR B1 to C2. Miura (2020) investigated the NICT-JLE corpus and extracted criterial pragmalinguistic features of L2 requestive speech acts, including errors (e.g., ellipses, phrases with topic-comment structures, declarative statements without tense inflections and modal verbs) in a general purchasing task and a negotiation task from CEFR A1 to B1. Huang, Lin, and Gráf (2023) examined the development of discourse markers such as *well*, *you know*, and *like* in L2 speech and extracted criterial features from the fluency-rated LINDSEI corpus across CEFR A2 to C1.

O'Keefe & Mark (2017) reviewed learner corpus-based English Profile Projects and CF studies and pointed out the significance of CF research on pragmatic development (i.e., "how a learner can use a syntactic form pragmatically and how they are able to skillfully play with a form for greater subtlety of meaning or focus" p. 479). Furthermore, although some previous studies in cognitive pragmatics have started to investigate construction-illocutionary force relationships in L1 CESCs, the same relationships in L2 CESCs across proficiency levels remain unexplored. Thus, the current study investigates which construction-illocutionary force relationships (the syntax-semantics-pragmatics interface) are criterial across adjacent CEFR levels in two Material World CESCs. Specifically, it focuses on the Processes schema constructions and the Object-schema constructions, both of which are infrequent and semantically peripheral. These constructions are assumed to be "pragmatically grounded and developed through social action" and to impose a high cognitive load on L2 learners, since performing the illocutionary forces requires metonymic cognitive shifts.

2.5 Research Questions

To investigate construction-illocutionary force relationships in L2 CESCs across proficiency levels, the current study addresses the following research questions (RQs):

RQ1. Which construction-illocutionary force relationships in TAM-specified Processes schema constructions are criterial in spoken interaction across adjacent CEFR levels?

RQ2. Which construction-illocutionary force relationships in TAM-specified Object-motion schema constructions are criterial in spoken interaction across adjacent CEFR levels?

3. Methods

3.1 Corpus

To identify L2 learners' illocutionary forces of canonical event schema constructions in spoken dialogues, the corpus used in the current study was an L2 learners' multimodal spoken dialogue (SD) component of the International Corpus Network of Asian Learners of English (ICANLE) (Ishikawa, 2023). The corpus includes 425 participants (including 20 native speakers), each producing four task modules (4250 files in total, Tokens = 779,621; Types = 9,517; Guiraud's R = 10.78, as calculated by CasualConc, Version 4.10, 2026) (Imao, 2026).

Among four tasks (Introduction, Part-time Job Task Set, Non-smoking Task Set, and Reflection), the current study focused on the two highly interactive Task Sets (Part-time Job Task Set and Non-smoking Task Set), each comprising a Roleplay and a Following Q&As, to address two research questions. It then focused on L2 learners' data only. Tables 4 (Spoken interaction tasks) and 5 (CEFR levels) summarise two versions of the descriptive statistics of the selected corpus.

	N	Tokens	Types	Ave Tokens (SD)	Ave Types (SD)	R
1.PTJ	405	75,208	2,407	185.70 (86.69)	75.45 (26.84)	8.78
2.PTJ Q&A	405	72,600	2,102	179.26 (92.68)	74.97 (28.33)	7.80
3.NS	405	81,486	3,004	201.20 (109.87)	86.55 (33.32)	10.52
4.NS Q&A	405	71,591	2,837	176.77 (90.56)	79.67 (27.97)	10.60

Notes. PTJ RP = Part-Time Job Roleplay; PTJ Q&A = Part-Time Job Q&A; NS RP = Non-smoking Roleplay; NS Q&A = Non-smoking Q&A; R = Guiraud's R

Table 4: Descriptive statistics of the selected corpus (Spok -en interaction tasks)

	N	Tokens	Types	Ave Tokens (SD)	Ave Types (SD)	R
A2	264	37,979	1,730	143.86 (66.92)	61.20 (20.70)	8.88
B1_1	356	64,831	2,648	182.11 (91.23)	76.02 (26.45)	10.40
B1_2	692	127,711	3,822	184.55 (90.77)	80.51 (27.71)	10.69
B2+	308	70,364	2,956	228.45 (114.32)	95.13 (34.15)	11.14

Notes. PTJ RP = Part-Time Job Roleplay; PTJ Q&A = Part-Time Job Q&A; NS RP = Non-smoking Roleplay; NS Q&A = Non-smoking Q&A; R = Guiraud's R

Table 5: Descriptive statistics of the selected corpus (CEFR Levels)

3.2 Data Analysis

To address two RQs 1 & 2, (1) all examples related to a prototypical verbs GET in the Processes construction and GO in the Object-motion schema construction were extracted from the selected SD component of ICNALE; (2) all the examples were manually coded in terms of the three perspectives: (a) two CESCs (i.e., the Processes schema construction and the Object-motion schema construction), (b) six TAM markers (i.e., Present, Past, Progressive, Perfective, Modality, and Semi-modality) and (c) thirty one illocutionary forces (e.g., expressing probability); (3) the Chi-squared test and standardized residuals first identified which TAM features occurred more or less significantly than expected across the two TAM specified constructions. The constructions and illocutionary force layers were analysed separately because combining them would produce excessively sparse cell frequencies; (4) the Random Forest (RF) classified the learner texts into CEFR levels based on the frequency of TAM-specified constructions, and identified candidate criterial features among the TAM markers using the Gini-based importance; (5) the Bayesian Additive Regression Trees (BART) model profiled each CEFR level's TAM-specified constructions based on posterior variable-inclusion proportions, and confirmed the reliability of the candidate features using 95% credible intervals; (6) the construction-illocutionary force relation-ships in the criterial TAM-specified constructions were quantified using an association measure, ΔP; and finally (7) frequent and strong construction-illocutionary force relationships in the criterial TAM-specified constructions across CEFR levels were identified from multimodal (video-based) data, with CEFR levels distinguished in terms of the degree of metonymic inferencing. All statistical analyses were conducted in R (Version 4.6.1, R Core Team, 2026) within the RStudio environment (Version 2026.08+195, Posit team, 2026).

4. Results

4.1 Construction-Illocutionary Force Relation-ships in the Processes Schema Constructions

4.1.1 TAM-Specified Constructions and CEFR Levels

First, to address RQ 1, I observed natural occurrences of the Processes schema constructions in the selected SD component of the ICNALE learner corpus, distinguishing six TAM markers (Present, Past, Progressive, Perfective, Modality, and Semi-Modality). Table 7 summarises the absolute and relative frequencies of the TAM-specified Processes schema constructions.

Second, the Chi-squared test statistically confirmed the relationships between TAM-specified constructions and CEFR levels.The results showed no statistical significances between them (χ 2(9, N = 59) = 6.40, p = .699; Fisher's exact test p =. 762, Cramer's V = .190). Because the overall Chi-squared test did not reach statistical significance, I examined standardised residuals to explore potential differences across cells in Table 7. However, Figure 2 also shows no remarkable differences between cells in the standardised residuals heatmap, as all residuals are below the threshold level (all |residuals| < 1.96).

TAM-specified constructions	A2	B1_1	B1_2	B2+
Processes/SVC (GET) +Present	7.90 (3)	9.25 (6)	7.83 (10)	15.63 (11)
Processes/SVC (GET) +Past	2.63 (1)	1.54 (1)	3.92 (5)	4.26 (3)
Processes/SVC (GET) +Progressive	5.27 (2)	1.54 (1)	5.48 (7)	2.84 (2)
Processes/SVC (GET) +Modality	5.27 (2)	0.00 (0)	2.35 (3)	2.84 (2)
Total	21.06 (8)	12.34 (8)	19.58 (25)	25.58 (18)

Note. Numerals are relative (normalized per 100,000 words) frequency; Numerals in parentheses are absolute (raw) frequency

Table 7: Frequencies of the TAM-specified constructions across CEFR Levels from A2 to B2+

Although none of the standardised residuals reached conventional significance (|z| <1.96), a level-specific tendency emerged: Modality was relatively more frequent than expected at the A2 level (z = 1.24), Present at the B1_1 level (z = 1.47), Progressive at the B1_2 level (z = 1.25), and Present again at the B2+ level (z = 1.04). Notably, all A2-level examples of Modality involved the modal verb *will* (e.g., *my friends* ***will get*** *nervous*).

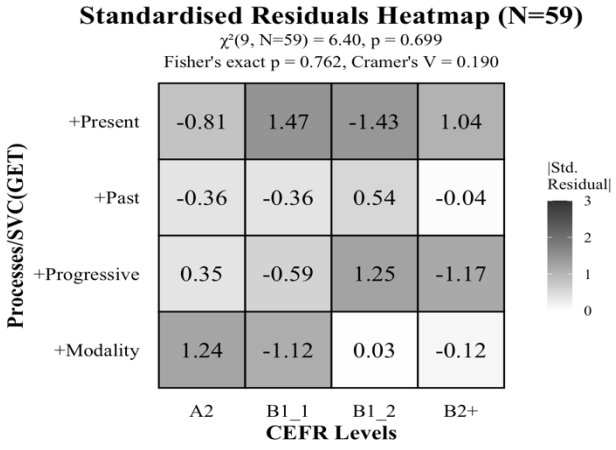

Figure 2: Standardized residuals heatmap (Processes/SVC (GET) constructions and CEFR levels)

An exploratory Random Forest classification was conducted using participant-level data (N = 42, five TAM-specified features) to examine which constructions clearly distinguished CEFR levels. Given the limited sample size, this analysis is exploratory. The Random Forest (1,000 trees, mtry = 2) achieved an out-of-bag (OOB accuracy of 42.90%, which did not exceed the majority-class baseline

(45.2%, always predicting B1_2), suggesting that the model's discriminatory power was limited given the available data. Nonetheless, the importance ranking still points to Present as the most important criterial feature, followed by Progressive, Past, and Modality.

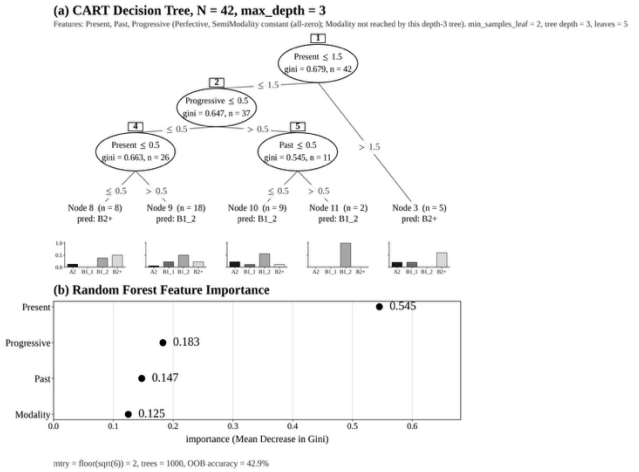

Figure 3: Random forest plot (Processes/SVC (GET) constructions and CEFR levels)

BART's posterior variable-inclusion proportions (Figure 4) profiled each CEFR level's TAM-specified construc tions: Present ranked highest for B1_1, B1_2 and B2+, co-occurring with Progressive for B1_2 and B2+, Past for B1_2 and Modality for A2 and B1_1. This was confirmed with 95% credible intervals, the reliability of the Random Forest's candidate features, and further identified Modality as an additional feature characterizing A2 and B1_1.

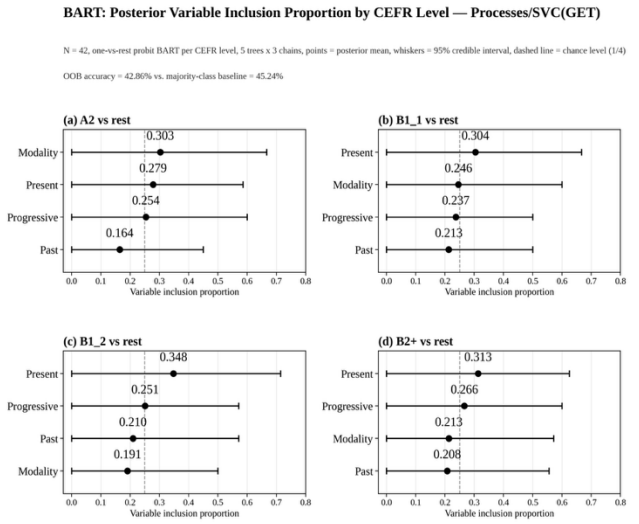

Figure 4: Bayesian additive regression tree (BART) plot (Processes/SVC (GET) constructions and CEFR levels)

4.1.2 TAM-Specified Constructions and Illocutionary Forces

The construction-illocutionary force relationships in the ΔP scores between the TAM-specified Processes/SVC (GET) schema constructions (Material World) in SD of ICNALE, which remained underexamined, and illocutionary forces were examined and are shown in Table 6.

L: CESCs	R: IFs	LR	RL	Freq
1.Processes/SVC(GET) +Present	Reporting	0.57	0.48	15
2.Processes/SVC(GET) +Progressive	Probability	0.52	0.98	12
3.Processes/SVC(GET) +Past	Reporting	0.36	0.87	9
4.Processes/SVC(GET) +Present	Probability	0.17	0.13	5
4.Processes/SVC(GET) +Modality	Probability	0.21	0.00	5
5.Processes/SVC(GET) +Present	Dissatisfaction	0.70	0.10	3

Notes. ΔP indicates the strengths and directions of illocutionary force salience from 0.00 to 1.00; LR shows left-to-right; RL shows right-to-left.

Table 6: Top 5 construction-illocutionary force relationships in the ΔP scores between the TAM-specified Processes/SVC (GET) schema constructions and illocutionary forces.

Top three examples of strong construction-illocutionary force relationships identified in the data are in (7) to (9) as follows:

(7) Processes/SVC(GET) Present →Reporting
(Freq 15, LR: 0.57)
e.g., *they get anxious* because *they just get addicted* to smoke. (ICNALE, B2+)

(8) Processes/SVC(GET) Progressive→Probability
(Freq 12, LR: 0.52)
e.g., *my friend is getting more serious* about that (ICNALE, B1_2)

(9) Processes/SVC(GET) Past→ Reporting
(Freq 9, LR: 0.36)
e.g., But *he got so nauseous* that we had to leave immediately, (ICNALE, B1_2)

In (7), both *they get anxious* and *they just get addicted* can be classified as the Processes/SVC schema construction with the present tense and with negative adjectives [DESCRIBING A PRESENT CHANGE OF STATE FOR REPORTING A PRESENT NEGATIVE MENTAL STA TE]. In (8), *my friend is getting more serious* can be analysed as Processes/SVC schema construction with the present progressive aspect [DESCRIBING A GRADUAL ONGOING CHANGE FOR EXPRESSING THE SPEAK ER'S SUBJECTIVE ASSESSMENT]. In (9), *he got so nauseous* can be construed as the Processes/SVC schema construction with the past tense [DESCRIBING A PAST CHANGE OF STATE FOR REPORTING A PAST NEGA TIVE MENTAL STATE].

4.2 Construction-Illocutionary Force Relationships in the Object-motion Schema Constructions

Similarly, to address RQ 2, I observed natural occurrences of the Object-motion schema constructions in the selected SD component of the ICNALE learner corpus, distinguishing six TAM markers (Present, Past, Progressive, Perfective, Modality, and Semi-Modality). Absolute and relative frequencies of the TAM-specified

Object-motion schema constructions are summarised in Table 8.

TAM-Specified CESCs	A2	B1_1	B1_2	B2+
Object-motion /SV(GO) +Present	0.00 (0)	1.54 (1)	2.35 (3)	1.42 (1)
Object-motion /SV(GO) +Past	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
Object-motion /SV(GO) +Progressive	0.00 (0)	1.54 (1)	3.92 (5)	1.42 (1)
Object-motion /SV(GO) +Modality	0.00 (0)	0.00 (0)	2.35 (3)	1.42 (1)
Total	0.00 (0)	3.08 (2)	8.61 (11)	4.26 (3)

Note. Numerals are relative (normalized per 100,000 words) frequency; Numerals in parentheses are absolute (raw) frequency.

Table 8: Frequencies of the TAM-specified Object-motion /SV(GO) constructions across CEFR Levels from A2 to B2+

The Chi-squared test statistically confirmed the relationships between TAM-specified constructions and CEFR levels. The results showed no statistical significances between them (χ 2(4, N = 16) = 0.99, p = .911; Fisher's exact test p = 1.000, Cramer's V = .176). As the overall Chi-squared test was not statistically significant, I examined standardised residuals to explore potential differenes across cells in Table 8. However, Figure 2 also shows no remarkable differences between cells in the standardised residuals heatmap, as all residuals are below the threshold level (all |residuals| < 1.96).

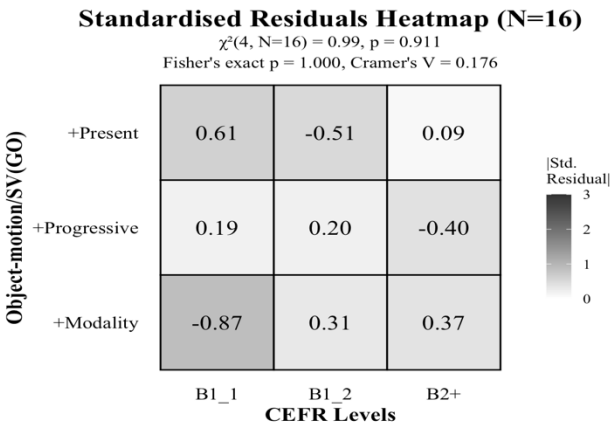

Figure 5: Standardized residuals heatmap (Object-motion/SVC (GO) constructions and CEFR levels)

An exploratory Random Forest classification was conducted using participant-level data (N=10, three TAM-specified features: Present, Progressive, and Modality) to examine which constructions clearly distinguished CEFR levels. Given the limited sample size, the analysis is exploratory (Figure 6). The Random Forest (1,000 trees, mtry = 1) achieved an out-of-bag (OOB accuracy of 40.00%, which did not exceed the majority-class baseline (60.00%, always predicting B1_2), suggesting that the model's discriminatory power was limited given the available data. Nonetheless, the importance ranking still points to Modality and Present as relatively more informative features; unlike the Processes/SVC (GET) construction, no clear level-specific separation emerged (most terminal nodes default to the majority class, B1_2).

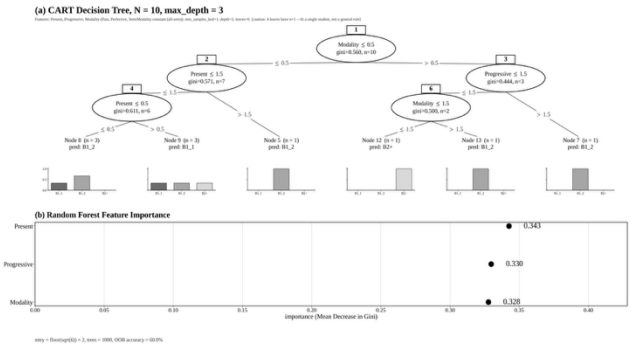

Figure 6: Random forest plot (Object-motion/SV (GO) constructions and CEFR levels)

BART's posterior variable-inclusion proportions (Figure 7) showed no clear level-specific patterns for this construction: all three features (Modality, Progressive, and Present) had posterior means clustered tightly around the uninformative baseline (1/3) with wide, heavily overlapping 95% credible intervals across all three CEFR levels, consistent with the Random Forest's limited discriminatory power given the small sample size (N = 10).

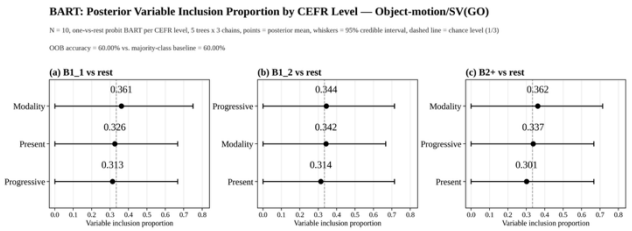

Figure 7: Bayesian additive regression tree (BART) plot (Object-motion/SV (GO) constructions and CEFR levels

As point estimates range only from .301 to .362, while the corresponding 95% credible intervals are wide (0.00-0.75) and overlapping, no remarkably significant differences were observed among the Object-motion constructions. This may be due to the imbalanced distributions of learners' utterances across CEFR levels, the absence of instances in A2 learners' utterances, and the overall low frequency of the Object-motion/SV(GO) constructions. Thus, because no candidate criterial features emerged among TAM-specified Object-motion/SV(GO) constructions, I did not calculate ΔP scores for the Object-motion/SV(GO) schema constructions, and I did not examine this construction type further through qualitative interpretation.

5. Discussion

Using the Random Forest and BART classifications, the present study investigated whether two grounded, TAM-specified Material World canonical event schema constructions (i.e., the Processes schema construction and the Object-motion schema construction, which are infrequent

and semantically peripheral in native data) are criterial across CEFR levels. Although no clear candidate criterial features were found through the Chi-squared tests and adjusted residuals, the results of the Random Forest (RF) and Bayesian Additive Regression Trees (BART) analyses implied that three TAM-specified Processes schema constructions were candidate criterial features with limited accuracy: (a) present progressive with comparatives [DESCRIBING A GRADUAL ONGOING CHANGE FOR EXPRESSING THE SPEAKER'S SUBJECTIVE ASSESSMENT] (B1 _2); (b) with past tense and negative adjectives [DESCRIBING A PAST CHANGE OF STATE FOR REPORTING A PAST NEGATIVE MENTAL STATE] (B1_2); (c) with the present tense and negative adjectives [DESCRIBING A PRESENT CHANGE OF STATE FOR REPORTING A PRESENT NEGATIVE MENTAL STATE] (B2+). Regarding Object-motion schema constructions, no critical findings emerged.

All three constructions are basic verb GET-based SVC sentence patterns *with a quietly, gradually shifting meaning* (i.e., ONGOING CHANGE, CHANGE OF STATE) and *metonymically inferred illocutionary forces* (i.e., EXPRESSING THE SPEAKER'S SUBJECTIVE ASSESSMENT, REPORTING A PRESENT (PAST) NEGATIVE MENTAL STATE). Specifically, the semantic schema and pragmatic force of these constructions seem difficult for L2 learners to explicitly or implicitly recognise and use during acquisition. Additionally, because the Process schema constructions and the Object-motion constructions are infrequent and semantically peripheral in native data, L2 learners are not frequently exposed to them in their L2 learning environments. In fact, the forms of the three TAM-specified constructions seem easy to acquire (i.e., basic verb GET-based SVC sentence patterns), while acquiring their semantic schema and pragmatic force takes considerable time and effort. Experientially, the three constructions appear to be valid candidate criterial features in L2 learners' interlanguage development at the syntax-semantics-pragmatics interface beyond B1_2 or intermediate to upper CEFR levels. Empirically, however, the statistical discriminatory power in the current study is small and needs to be deliberately re-examined with larger L2 spoken interaction learner corpora in the same way.

Pedagogically, if these constructions are confirmed to be reliable criterial features, material designers will be able to design lessons from three-layered perspective of the interface syntax-semantics-pragmatics, focusing on a single basic verb GET: for example, (a) grammatical target: present progressive with comparatives; (b) semantic target: [DESCRIBING A GRADUAL ON GOING CHANGE; and (c) pragmatic target: EXPRESSING THE SPEAK ER'S SUBJECTIVE ASSESSMENT] (B1_2). Additionally, teachers also can assess L2 learners' proficiency development in spoken interaction activities from the three-layered perspective in both L2 grammar instruction and pragmatics instruction.

6. Conclusion

The current study explored whether two grounded, TAM-specified Material World canonical event schema constructions, which are infrequent and semantically peripheral in native data, are criterial across CEFR levels.

For further research, other frequent constructions among 13 CESCs (see Table 1) should be examined from the same perspective of the three-layered interface between syntax, semantics and pragmatics. Experientially and empirically, infrequent, semantically peripheral and pragmatically inferred constructions are candidate criterial features between intermediate and higher levels even though their forms are simple; on the other hand, frequent constructions with simple semantic schema and less inferred pragmatic forces (e.g., the States schema constructions: *This is true*) are assumed to be candidate criterial features at lower levels. Thus, taking a three-layered perspective (syntax-semantics-pragmatics) on sentence-level constructions in criterial features research could lead us to a more balanced instruction of L2 grammar and pragmatics in the action-oriented approaches to L2 pedagogy.

Acknowledgements

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Numbers: JP24K04092.

References

Adolphs, S. (2008). *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam: John Benjamins.

Aijmer, K. (1996). *Conversational Routines in English: Convention and Creativity*. London, UK: Longman.

Aijmer, K. (2013). *Understanding Pragmatic Markers: A Variational Pragmatic Approach*. Edinburgh, UK: Edinburgh University Press.

Aijmer, K. & Rühlemann, C. (2015). *Corpus Pragmatics: A Handbook*. Cambridge, UK: Cambridge University Press.

Alston, W. P. (2000). *Illocutionary Acts and Sentence Meaning*. Ithaca, NY: Cornell University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* London, UK: Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (2021). *Grammar of Spoken and Written English.* Amsterdam: John Benjamins.

Carter, R. & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge, UK: Cambridge University Press.

Chen, Y.H. & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays. *Applied Linguistics, 37*(6):849−880.

Dunn, J. (2022). *Natural Language Processing for Corpus Linguistics*. Cambridge, UK: Cambridge University Press.

Fillmore, C. J., Johnson, C. R., & Petruck, M. R. (2003). Background to Framenet. *International Journal of Lexicography, 16*(3): 235−250.

Galaczi, E.D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics, 35*(5):553 −574.

Hawkins, J.A. & Buttery, P. (2010). Criterial features in learner corpora: theory and illustrations. *English Profile Journals, 1*(1):1−23.

Hawkins, J.A. & Filipović, L. (2012). *English Profile Studies 1: Criterial Features in L2 English.* Cambridge, UK: Cambridge University Press.

Huang, L., Lin, Y., & Gráf, T. (2023). Development of the use of discourse markers across different fluency levels of CEFR: A learner corpus analysis. *Pragmatics, 33*(1): 49−77.

Huddleston, R., & Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.

Imao, Y. (2026). *CasualConc* (Version 4.10) [Computer software]. Retrieved from https://sites.google.com/site /casualconcj/

Ishikawa, S. (2023). *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian's Learners' L2 English*. Abingdon, UK: Routledge.

Jucker, A.H. & Taavitsainen, I. (2013). *English Historical. Pragmatics*.Edinburgh, UK: Edinburgh University Press.

Kang, O. & Wang, L. (2014). Impact of different task types on candidates' speaking performances and interactive features that distinguish between CEFR levels. *Research Notes, 57*:40−49.

Khushik, G.A. & Huhta, A. (2020). Investigating syntactic complexity in EFL learners' writing across the Common European Framework of Reference levels A1, A2, and B1. *Applied Linguistics, 41*(4):506−532.

Langacker, R.W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford, UK: Oxford University Press.

Langacker, R.W. (2009). Cognitive grammar. In D. Sandra, J-O. Östman, & J. Verschueren. (2009). *Cognition and Pragmatics*. (pp. 78-85). Amsterdam: John Benjamins.

Littlemore, J., Krennmayr, T., Turner, J., & Turner, S. (2014). An investigation into metaphor use at different levels of second language writing. *Applied Linguistics, 35*(2): 117−144.

Mazzone, M. (2019). *Cognitive Pragmatics: Mind Reading, Inferences, Consciousness*. DE Gruyter/Mouton.

Miura, A. (2020). Criterial pragmalinguistic features of Requestive speech acts produced by Japanese learners of English. *Learner Corpus Studies in Asia and the World 4*, 1−23. Kobe University Repository.

Notohara, Y. (2016). *Tenkeiteki na Kobun to Doshi* [Canonical constructions and verbs in Spoken English]. In Y. Yanase & T. Nishihara (Eds.), *Kotoba de Hirogaru Chisei. to Kansei no Sekai* [Sense and sensibility through languages]. Keisuisha.

Notohara, Y. (2021). Type/token frequency effects. in L2. learners' canonical construction development. *Doshisha Studies in English, 102*, 141–178.

Notohara, Y. (2023a). Exploring illocutionary force of tense-aspect-modality (TAM) specified canonical constructions: A collostructional approach. *The 18th International Pragmatics Conference (IPC)*, Brussels, Belgium.

Notohara,Y.(2023b). Conventionalised illocutionary scenarios and figurative thought patterns from the psychological schema constructions in spoken English. *The 16th International Cognitive Linguistics Conference. (ICLC)*. Düsseldorf, Germany.

Notohara, Y. (2025). Conventionalised illocutionary forces. of the canonical event schema constructions in Spoken English: A collostructional approach. *Journal of Corpus-based Lexicography Studies, 7*, 45−66.

O'Keefe, A. & Mark, G. (2017). The English grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, *22*(4): 457−489.

Panther, K-U.(2022). *Introduction to Cognitive Pragmatics*. Amsterdam: John Benjamins.

Panther,K-U., & Thornburg, L. (1998). A cognitive approach to inferencing in conversation. *Journal of Pragmatics*, *30*(6):755−769.

Panther, K-U., & Thornburg, L. L. (Eds.). (2003). *Metonymy and Pragmatic Inferencing (Vol.113).* Amsterdam: John Benjamins.

Panther, K-U., & Thornburg, L. L. (2017). The role of inferencing in the interpretation of two expressive speech act constructions. In F.J. Ruiz de Mendoza Ibáñez, A.L. Oyón, & P. Pérez-Sobrino (Eds.), *Constructing Families of Constructions: Analytical Perspectives and Theoretical Challenges*. Amsterdam: John Benjamins, pp. 109−134.

Posit Software. (2026). *RStudio* (Version 08.1+195) [Computer software]. Retrieved from https://docs.posit. co/ide/user/#rstudio-ide-oss-downloads.

Radden, G. & Dirven, R. (2007). *Cognitive English Grammar*. Amsterdam: John Benjamins.

Römer, U. & Verger, C.M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition, 41*:1089−1110.

Rühlemann, C. (2019). *Corpus Linguistics for Pragmatics: A guide for research*. Abingdon, UK: Routledge.

Ruytenbeek, N., Ostashchenko, E., Kissine, M. (2017). Indirect request processing, sentence types and illocutionary forces. *Journal of Pragmatics, 119*:46−62.

Schmid, H.J. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment.* Oxford, UK: Oxford University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition.* London, UK: Blackwell.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition.* 2nd Edition. London, UK: Blackwell.

Stefanowitsch, A. (2003). A construction-based approach to indirect speech acts. In K-U, Panther & L.L. Thornburg. (Eds.), (2003). *Metonymy and Pragmatic Inferencing*. Amsterdam: John Benjamins, pp. 105−126.

Thornburg, L., & Panther, K. (1997). Speech act metonymies. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*:205−222.

van Ek, J.A. & Trim, J.L.M. (1991). *Threshold 1990*. Cambridge, UK: Cambridge University Press.

Weisser, M. (2015). Speech act annotation. In K. Aijmer & C. Rühlemann. (Eds.), *Corpus Pragmatics: A handbook*. Cambridge, UK: Cambridge University Press, pp. 84−113.

Weisser, M. (2020). Speech acts in corpus pragmatics: Making the case for an extended taxonomy. *The International Journal of Corpus Linguistics*, *25*(4): 400−425.