

DOI: 10.14746/linpo.2024.66.1.1

The use of discourse markers in conversation: The case of “myślę” – “I think” in Polish Career Clean Coaching

Konrad Juszczuk¹ & Barbara Konat²

¹ Adam Mickiewicz University in Poznań
juszczuk@amu.edu.pl | ORCID: 0000-0002-4769-0155

² Adam Mickiewicz University in Poznań
bkonat@amu.edu.pl | ORCID: 0000-0003-2370-4636

Abstract

People find it difficult to talk about abstract concepts such as their future career plans. The discourse marker “I think” is used to express uncertainty and soften the force of a proposition. The objective of this paper is to demonstrate how specific context and bounded topic of conversation influence the use of the discourse marker “I think” in Polish (*myślę*) as a structuring discourse marker. A total of 23 participants were recorded during Career Clean Coaching sessions and the usage of “I think” in the resulting corpus was compared to other Polish conversational corpora. The qualitative study of the semantics of “I think” indicates that this discourse marker serves a hedging function. The quantitative analysis of the frequency of the discourse marker “I think” in Polish conversational corpora and in the Career Clean Coaching corpus reveals significant differences. The results indicate that pragmatic factors, namely the topic, context and protocol of the conversation, exert a strong influence on the usage of “I think”.

Keywords: discourse markers, conversations, corpus study, clean coaching, pragmatics

1. Introduction

This paper presents the Career Clean Coaching corpus and compares it with two other Polish corpora: goal-oriented dialogue and a sample of natural, everyday conversations. The aim is to explore how the context of conversations influences the frequency of “I think” as a marker of uncertainty. To achieve this, we provide a detailed analysis of the Career Clean Coaching sessions and corpus, as well as a semantic distinction of the five basic senses of “I think” in Polish, which were annotated in the Career Clean Coaching corpus. Our initial study of coaching sessions has suggested that “I think” is

a keyword with high frequency, among other cognitive verbs (Konat & Juszczuk 2015). This initial finding motivates the analysis of “I think” as a discourse marker in the current paper.

This paper is structured as follows: Section 2 presents related work concerning the role and functions of discourse markers in English and Polish in different communicative contexts. Section 3 provides a detailed description of the data used in the study. In this section, the features of coaching as conversation are presented and the Clean Coaching protocol is introduced, along with the introduction of the Career Clean Coaching (CCC) corpus as a material. The fourth section presents the results of Studies 1-3.

Firstly, it is predicted that the participants (coachees) of Career Clean Coaching sessions will use the discourse marker “I think” mainly to soften the strength of their claims. The first hypothesis is that among all possible senses of “I think” in Polish, the most frequent will be the one associated with the function of a hedge. This is illustrated in our Study 1: Five senses of “I think” in Career Clean Coaching.

Secondly, we posit that the frequency of usage of “I think” is not solely a characteristic of an individual speaker but is also shaped by the form of conversation. Our second hypothesis is that the specific protocol, as well as abstract topics of Career Clean Coaching sessions, will result in a high frequency of “I think”. To test this, we compared the Career Clean Coaching corpus with the corpora of other Polish conversations in Study 2: The frequency of “I think” in Career Clean Coaching.

Thirdly, we posit that although the usage of “I think” varies between speakers, the context, topic and specific protocol exert a stronger influence. This is explored in Study 3: Speaker differences in Career Clean Coaching, where inter-speaker variability is examined.

2. Related work

The research presented here is informed by theories of cognitive linguistics, conversation analysis and studies on discourse markers. The functions of discourse markers may vary depending on the communicative context in which they are used (Tay 2013: 164). They may indicate solidarity, rapport, and in-group membership (Cutting 2000), contribute to the perceived naturalness of conversation (McCarthy 1998), and convey a general sense of indirectness, often interpreted as being more polite (or less intrusive) in conversation (Brown & Levinson 1987). Tay’s analyses of psychotherapy extracts suggest that discourse markers may be directly motivated by metaphor. In some cases, they may signal upcoming metaphorical expressions. Although Tay mentions that his analysis is not validated on large corpora, he also gives reasons why the use of discourse markers and metaphors may be interrelated. Metaphors are employed in conversation to convey evaluations, attitudes, values, perspectives, and beliefs (Cameron & Maslen 2010). Additionally, they are utilized when individuals “struggle to find words to capture difficult to describe sensations” (McMullen & Conway 1996: 252). Discourse markers such as “you know” or utterance-final “right” occur at significant information state transitions, where the objective of the speaker is to reduce an initial asymmetry with regard to some body

of knowledge held by the speaker and hearer (Tay 2013: 161). In this paper, we argue that “I think” in coaching sessions serves similar functions.

The data-driven approach, similar to conversation analysis of psychotherapeutic sessions (Ferrara 1994; Pawelczyk 2011; Tay 2013), is employed in this research. The analysis of coaching sessions as dialogues is framed in conversation analysis, as its scope is a professional discourse, and it relies on recorded interactions (sessions) of naturally occurring speech. The discourse marker “you know” has been identified as a facilitator of intimacy between clients and psychotherapists. This is evidenced by the observation that it “mitigates the threat posed to the client in verbalizing highly intimate information” (Pawelczyk 2011: 154). The marker is often used to precede a potentially threatening or traumatic thought or idea that is about to be revealed by the client. (Pawelczyk 2011: 151). This is why discourse markers such as “you know” often signal self-disclosure of the patient in psychotherapy.

Our study is also closely related to work on discourse markers of uncertainty in English and Polish. In this area, studies of “I think” have shown its versatile functions. In her study of rape trial discourse, Ponterotto (2014) included the phrase “I think” in the category of hedges, which are used by speakers when reporting difficult and traumatic experiences. Mullan (2010) posits that the extensive use of “I think” is a strategy for dealing with uncertainty because it mitigates or softens the force of a proposition.

The reasons for downtoning a proposition can include being unsure as to the truth or reliability of a statement, wishing to claim or disclaim responsibility for a proposition, and mitigating or softening a statement out of politeness or consideration for the addressee (Mullan 2010: 59).

In English conversations, “I think” is treated as an indicator of taking an epistemic stance (Kärkkäinen 2003). Spoken corpora have been employed to investigate the frequency of the use of “I think” as an indicator of uncertainty (Brezina 2013). The use of cognition verbs in the first-person singular may be regarded as a distinctive feature of conversational interactions. Helasvuo (2014) observed that in Finnish conversations, the verb “*ajatella*” – “to think” – appeared among the five most frequent cognition verbs, with a total of 132 instances. Among the instances, as much as 90 (68%) were used in the first-person singular form. Consequently, this paper focuses on the exact word form “*myślę*” – “I think”, in which speakers are referring to their own cognitive states.

One of the cases where the pragmatic factor – a specific context of usage – influences the function of “I think” is the use of English either as a first language or second language. Baumgarten & House (2010) analysed the functional profiles of “I think” and “I don’t know” in conversations between native speakers and in English as a *lingua franca* interaction. Biber (2012) provided quantitative evidence for the influence of the discourse register on the frequency of lexemes and structure, demonstrating how conversations differ from information writing. In this paper, we conduct a comparison between a balanced corpus, which is mostly written, and our coaching session corpus, confirming Biber’s observation. We also provide evidence that pragmatic factors of conversation, such as setting, topic, and speakers, can strongly influence the lexical choices of the speaker.

Polish is a null-subject language. Some researchers (see, for example, Posio 2014) have argued that the presence or absence of an overt subject in the conversational use of “I think” should be considered in analyses. However, in the current study, we did not analyse the presence of an overt subject, as it was very rare (only 10 instances), which did not allow for generalisations. The position of “I think” in the phrase, either initial (Kaltenböck 2009) or final (Sato 2017), has been studied as a meaningful factor. However, this aspect is not considered in the current study.

The study presented here demonstrates that the usage of “I think” in natural, spoken communication, where the owner of the meaning of “I think” is the actual speaker, differs from that in official or written text. The method employed in our study is corpus analysis of language data, which is sometimes referred to as “a combination of interactional sociolinguistics and corpus linguistics” (Baker 2010). Finally, this paper aims to contribute to the study of the pragmatics of Polish spoken language. A study conducted on a sample of 100 articles from the *Journal of Pragmatics* revealed that only one article in the sample addressed Polish as a subject of study. Polish was identified as one of 20 languages with a single publication, whereas English was studied in as many as 50% of the articles (Egbert et al. 2016).

A number of studies have examined the use of the discourse marker “I think” in different languages and contexts. A multimodal analysis of the discourse marker “I think” in Brazilian Portuguese was conducted on a sample of audio-video recorded interviews (Freitag et al. 2021). The use of “I think” and similar phrases was examined in task-based group conversations with native and non-native Swedish speakers (Tolvanen 2024). The translation of the phrase “I think” from English to Lithuanian and Hebrew was examined using the alignment model of phrase-based statistical machine translation, with manual treatment of the data (Oleskeviciene & Liebeskind 2021). In conclusion, these studies demonstrate the multifaceted and context-dependent nature of the discourse marker “I think”.

3. Material – Coaching Sessions

3.1. Coaching session

A coaching session is a conversation between a coach and a coachee, during which the coach employs a specific protocol and mirroring technique to encourage the coachee to self-disclose personal and professional experiences. Coaching is provided to employees in corporations, entrepreneurs, members of non-governmental organisations, and students in career advising offices. Individuals engage with a coach when they require assistance in developing self-esteem and self-confidence, as well as in enhancing their soft skills, including interpersonal communication, public speaking, work performance and work-life balance, time and team management, career change and other areas. The primary reason for selecting coaching is the perception that one’s performance and success in life do not align with the expectations of one’s supervisor or oneself.

In contrast to other forms of support, such as counselling, psychotherapy, mentoring or training, the coach does not typically provide advice to the coachee on how to change their attitude or performance. The individual undergoing coaching is referred to as a “client” rather than a “patient”. The coach is not a therapist, as the coach does not diagnose the client and does not provide them with solutions, advice, or treatment. Instead, the coach guides the coachee through the process of change, with the objective of instilling self-steering. The role of the coach is to provide support to the coachee in understanding the nature of the problem, identifying the desired outcome and the necessary resources, and facilitating the coachee’s willingness to change. The coachee is encouraged to provide a summary of their experience and to interpret the content of the session. This process enables the coachee to gain self-steering and independence in decision-making and action-taking. During the coaching session, the coachee is encouraged to explain their situation, identify possibilities, make decisions and take actions independently. Furthermore, coaching differs from psychotherapy in that the coach is focused on the coachee’s future, rather than their past experience. The coach also draws the coachee’s attention to their resources and helps them to find a solution that is suitable for their situation (Thorpe & Clifford 2003; Pawelczyk 2011; Sullivan & Rees 2008).

3.2. Coaching sessions model

In this section, we compare coaching with psychotherapy and everyday conversation. Since coaching is a form of providing support in conversation, we assume that it is similar to consultation or psychotherapy. To contrast coaching discourse with other kinds of conversation, we use Ferrara’s model for differentiating psychotherapy from ordinary conversation (Ferrara 1994). Ferrara considers regular conversation to be the unmarked form of discourse (Ferrara 1994). In regular conversation, both speakers contribute equally to the content and details. The way in which speakers exchange information depends on the specific needs of the individual and the type of information being shared. For example, friends or family members meet to share their experiences and to spend time together, so the aim, form and topic of conversation are not explicitly stated, and the form is less regulated. According to Ferrara’s model, there are seven dimensions which distinguish psychotherapy from conversation:

- 1. Parity**
- 2. Reciprocity**
- 3. Routine recurrence**
- 4. Bounded time**
- 5. Restricted topic**
- 6. Remuneration**
- 7. Regulatory responsibility**

It can be argued that a coaching session differs from a conversation in the same way as a psychotherapy session does (see Table 1). All seven features of psychotherapy discourse are present in coaching discourse. We propose the application of Ferrara’s model

to coaching, which allows us to present how a coaching session is similar to a psychotherapy session.

1. Parity: on this dimension, a coaching session is uneven, with a lower parity in comparison to everyday conversation. The client (coachee) agrees to be supported by the coach and allows the coach to conduct the conversation according to a certain protocol.

2. Reciprocity: This is absent in coaching, as only the coach inquires about the coachee's experiences and attitudes, rather than the converse.

3. Routine recurrence: Coaching sessions constitute a fixed element of the coaching process, occurring at regular intervals and for a pre-determined duration. These meetings are a recurring routine for both parties.

4. Bounded time: Coaching sessions typically span a duration of between 30 and 90 minutes.

5. Restricted topic: The **topic** of the coaching session may be selected by the client (coachee) or proposed by the sponsor (company funding the coaching). Depending on the coaching type, the topic may relate to management skills, decision-making and task delegation in business coaching or work-life balance, mental health and self-fulfilment in life coaching, for example.

6. Remuneration: Coaching differs from conversation in that it is a professional service provided by trained specialists. Coaches are remunerated either by the client or by a sponsor.

7. Regulatory responsibility: The coach is responsible for the form of discourse development since the coach is the one who opens the session, forms questions for the coachee and terminates the session. However, the coach is focused on the coachee's career and experience, and therefore the specific content of the session is dependent on the coachee. It is not permitted for the coach to introduce any details of their personal life or professional career.

In order to describe Career Clean Coaching sessions, we propose the development of a new model based on Ferrara's existing model. Our new model incorporates three additional dimensions: self-disclosure, mirroring, and protocol. The last two dimensions have different values in Clean Coaching and psychotherapy, thus constituting the distinctive features of Clean Coaching.

8. Self-disclosure: in psychotherapy, coaching or everyday conversations, self-disclosure is understood as "the process of deliberately revealing information about oneself that is significant and that would not normally be known by others" (Adler & Towne 1996: 358). However, as Pawelczyk observes, "in psychotherapy, unlike in other social contexts, it is the client who is granted extensive clinical and conversational space to disclose" (Pawelczyk 2011: 123). Similarly, the coaching session represents another social context in which the coach provides the client with a conversational space to disclose and freely discuss their experiences and expectations in life. In regular conversation, self-disclosure can be mutual and equal, with both parties sharing their experiences and exchanging news about their lives.

9. Mirroring is an important aspect of building rapport and occurs naturally and spontaneously in conversation (Chartrand & Bargh 1999). Typical mirroring is manifested through similar body posture, gestures, intonation, and phrases (Kipp 2010). Some

psychotherapists and coaches are trained in the practice of mirroring and employ it deliberately (Thorpe & Clifford 2003). Psychotherapists frequently repeat their patients' words with the intention of fostering their patients' self-expression and indirectly requesting further elaboration (Pawelczyk 2011: 163). Similarly to psychotherapists, coaches combine exact repetitions or paraphrases of their clients' words with questions to draw attention to salient issues of their clients' experiences. This encourages clients to share their experiences in the process of self-disclosure (Thorpe & Clifford 2003). Such verbal mirroring is also referred to as back-channelling or paraphrasing (Pawelczyk 2011).

10. Protocol: a set of rules and speech acts that are typically employed by dialogue participants. However, in contrast to regular conversation, where the protocol is not explicitly known, the protocol used in psychotherapy or coaching is maintained by the psychotherapist or coach, respectively. The specific protocol governing the conduct of psychotherapy or coaching sessions varies according to the therapy type and determines the general aim and type of support provided. The specific protocol is explicitly known to the coach but not to the coachee. The detailed protocol of coaching sessions on which our study was based is described in the next section (Clean Coaching Protocol).

Table 1: Comparison between conversation, psychotherapy session and Clean Coaching

Dimensions	Conversation	Psychotherapy	Clean Coaching
1. PARITY	high (dialogue partners are equal)	low (in favour of psychotherapist)	low (in favour of coach)
2. RECIPROCALITY	high (dialogue partners are equal)	suspended	suspended
3. ROUTINE RECURRENCE	no	yes	yes
4. BOUNDED TIME	no	yes	yes
5. RESTRICTED TOPIC	no	yes	yes
6. REMUNERATION	no	yes	yes
7. REGULATORY RESPONSIBILITY	dialogue partners are equal	psychotherapist	coach
8. SELF-DISCLOSURE	mutual and equal	patient's	coachee's
9. MIRRORING	natural and spontaneous	performed by trained psychotherapist in form of paraphrase	performed by trained coach in form of parrot-phrasing (exact wording)
10. PROTOCOL	does not need to be stated and known explicitly	depends on the type of therapy	Clean Language

Table 1 summarizes how the coaching session is different from regular conversation, but similar to the one between psychotherapist and patient in psychotherapy (Juszczak 2017).

3.3. The Clean Coaching Protocol

Clean Coaching is a type of coaching that employs a special set of questions and techniques, collectively known as Clean Language, which constitute the protocol for the session. Clean Language questions were developed by David Grove for psychotherapy in the late 1980s and subsequently established as a system by Lawley and Tompkins (2000). The protocol is currently employed worldwide in coaching, consultancy, conflict resolution, interviewing or interrogation, and business management and education (Sullivan & Rees 2008). In the current study, the Polish adaptation of Clean Language is utilised (Pieśkiewicz & Kołodkiewicz 2011). In this method, the coach selects questions from a limited set that facilitate the coachee in reflecting on their experience and developing metaphors of their experience. Metaphors are believed to facilitate discourse on complex and abstract topics, such as the future, development, or job-seeking (Sullivan & Rees 2008).

In Clean Language, questions are designed to be “clean”, which entails that the coach does not propose any metaphorical expression but repeats the coachee’s words without modification. This method of questioning is referred to as parrot-phrasing, in contrast to paraphrasing, where the coach modifies and reinterprets the coachee’s words. Parrot-phrasing results in numerous repetitions of the coachee’s phrases during the session. However, this approach simultaneously ensures that the coachee remains consistent and coherent in their responses, focusing on a specific topic.

The prescriptive protocol presented in Table 2 is reconstructed on the Clean Language handbook for practitioners (Sullivan & Rees 2008). The session commences with the Opening Question (also referred to as the Intention Question), which is typically of the following form: “What would you like to have happen?” or “When you think about your career, this is like what?”. Subsequently, the coach repeats the coachee’s words and requests further information using development questions, such as “Is there anything else about X?” or “What kind of X is that X?” where X represents the client’s previous statements. Once the coachee has developed a metaphor for their career, the next stage of the session involves the use of so-called attributes and location questions. The coach may inquire as to whether the concept in question has a shape or size, or alternatively, where it is located. These questions are designed to assist the coachee in identifying connections and associations between the concepts. As more metaphors emerge in the coachee’s words, the coach may pose the Relationships Question, for instance, “Is there a relationship between X and Y?”. At this stage, the coach may also introduce Sequence and Source Questions, which are of the following form: “What occurs immediately preceding X?” or “What occurs next after X?” to conclude the session. The final stage of the session may include the questions, “What is the first thing you know now?” and “What is the second thing you know now?”. This series of questions continues until the coachee responds to the question about the sixth thing the coachee knows as a result of the session. The question typically comprises the following elements: a conjunction (and), a pronoun (that, this, these, those), the coachee’s words, and a clean question, followed by the coachee’s words again. An example of a complete question is “And that eagle: what kind of eagle is that eagle?” or “And that rollercoaster: is there anything else about

that rollercoaster going fast?” or “And that surprise: does that surprise have a shape or a size?”. This general protocol of Clean Coaching was used to obtain the Career Clean Coaching corpus for the current study.

Table 2: Overview of the protocol of Clean Coaching session

Type of question	General form of the question
Intention Question (Open)	“What would you like to have happen?” “When you think about your career this is like what?”
Development Questions	“Is there anything else about X?” “What kind of X is that X?”
Attribute and Location Questions	“Does X have a shape or size?” “Where is X?”
Relationships Questions	“Is there a relationship between X and Y?”
Sequence and Source Questions	“What happens just before X?” “When X what happens next?”
Knowing Questions (Terminate)	“What is the first (second, third etc.) thing you know now?”

It can be argued that these dimensions of coaching sessions set them apart from everyday conversations. These differences in the pragmatic set up of the interaction influence the lexical choices of speakers. In particular, the frequency of “I think” as a discourse marker is worthy of further investigation.

3.4. Clean Career Coaching corpus

The Clean Career Coaching corpus comprises 23 dialogues concerning the future careers of participants. These dialogues were recorded during free coaching sessions on professional development and career planning. The sessions were advertised via university career offices and a project website. The coaching sessions were conducted according to the model presented in Section 3.2 and the set of Clean Questions presented in Section 3.3. The participants were between the ages of 25 and 35 and were currently enrolled in or had recently graduated from a Master’s or PhD program. This demographic may be expected to have concerns about their careers, as some had recently begun their professional lives, while others were still engaged in the job search process. Each subject participated in two coaching sessions with one of two hired coaches, who assisted them in defining career objectives and making decisions about their future careers. The coaching sessions were recorded with multiple cameras and microphones. None of the researchers were present in the recording room during the actual sessions, and the coach and coachee were afforded as much privacy as possible. Written consent for being recorded was obtained from all participants, and each was granted the right to withdraw at any time. Consequently, the participants were unaware of the actual aim of the research project,

which was a comprehensive study of multimodal communication in coaching (including the analysis of gestural metaphor). Furthermore, participants were afforded the opportunity to engage in discourse with the coach prior to and following each session, without being recorded, thus enabling them to address any concerns or receive additional feedback.

The coaches who recorded sessions for this study underwent training in accordance with the standards set forth by the International Coaching Federation (ICF) and were certified as Associate Certified Coaches by the ICF. This certification attests to their competence in supporting individuals in their personal and/or professional development. Additionally, they received training and certification as Clean Language coaches in the Polish version of Clean Language. The coaches who were recorded as part of the Career Clean Coaching corpus were hired and remunerated as part of a funded research project. Each coachee who was recorded as part of the Career Clean Coaching corpus participated in two sessions. The average duration of a session in the Career Clean Coaching corpus was 40 minutes. While the topic was limited to career-related matters, the coachees were permitted to select any aspect of their professional careers. This could include their current circumstances, future aspirations, or their general understanding of professional careers in the context of other activities in their lives. A portion of the recordings was transcribed using ELAN software (Sloetjes & Wittenburg 2008), and the utterances of the coachees were utilized in the subsequent analyses. The resulting Career Clean Coaching corpus, which was employed in this study, encompasses 104,914 words from 23 speakers (coachees), 18 of whom are female and 5 of whom are male.

4. Results

4.1. Study 1: The Five senses of “I think” in Career Clean Coaching

The verb “to think” (“myśle” in Polish) is highly polysemous in Polish, as it is in many other languages. A study utilizing dictionary-based annotation enriched with cognitive linguistics concepts of agentivity and conducted on the Polish written corpus revealed varying degrees of agentivity in five distinct senses of the verb “to think” (Kokorniak & Konat 2012). Increasing the sense of agentivity is a primary objective for both coaches and coachees, and the usage of “I think” may indicate whether the speaker perceives themselves as an agent or an experiencer. The objective is to identify the same set of five senses in the Career Clean Coaching corpus and to focus on the first-person singular form of the verb “I think”.

A manual annotation was conducted on all 300 instances of “I think” in the Career Clean Coaching corpus, following the methodology introduced in Kokorniak & Konat (2012). The taxonomy employed here adheres to a cognitive view of language, establishing a continuum between the subject as an agent and the subject as an experiencer. The subject as an agent is the most active, volitional, controlling and responsible, corresponding to Sense 5. In contrast, the subject as an experiencer is more passive, reflecting and re-membering (Sense 1). In order to demonstrate the distinction between the five senses, we propose an English equivalent for each example, which we believe is the closest to

the given facet of the meaning conveyed by the speaker. At the same time, it should be noted that in English, the word “think” can also be used in a word-by-word translation.

Firstly, the use of “I think” allows speakers to express mental processes in which they are not responsible (such as the process of remembering) and non-volitional, as exemplified in Example 1 (Sense 1). An indicator of this sense is the construction “I think about” – “myślę o”. An English equivalent of this meaning can be conveyed by the word “remember”.

Example 1:

*no taką że czuję ulgę w momencie nawet kiedy sobie **myślę** o tym że ja tej kłody nie mam na rękach*

(particle).such.that.feel.relief.in.moment.even.when.1SGthink.about.this.that.I.this.log.
not. have.on.hands

[I feel relieved in the moments **I think** that I no longer have this log on my hands.]

Secondly, the process of thinking may be described as more volitional, whereby the speaker is establishing relationships between different concepts in the process of imagination. This process is volitional, yet the speaker lacks complete control over it and is not solely responsible for the truth value of it, as evidenced by Example 2 (Sense 2). In English, this can be expressed with the verb “to imagine”.

Example 2:

*nie wiem czy się będę w niej sprawdzać po prostu sprawdzić czy jest faktycznie tak jak **myślę** tak zweryfikować.*

no.know.if.(reflexive).will.in.it.perform.just.simple.check.if.is.really.just. how.1SGthink.
like.verify

[I don’t know if I will perform well in it, just to check if this is really how **I think** it will be, just to verify.] or [I believe that the best way to proceed is to verify this information.]

Thirdly, when “I think” precedes the proposition, it is employed to diminish the assertion that follows, which aligns with Sense 3 in our taxonomy. This usage is frequently accompanied by the use of the word “że” (“that”) and generally corresponds to the English use of “I think” as a discourse marker used to hedge propositions, as shown in Example 3.

Example 3:

*ja **myślę** że to taki cel do którego gdzieś tam dążę to co chciałabym osiągnąć* I.think.
that.it.such.aim.to.which.somewhere.there. strive.this.what.would.acheive

[**I think** that this is the goal towards which I am striving, something I am striving to achieve.]

Fourthly, more volitional usage is related to the expression informing about speakers’ plans. In Polish, “to think” may be used to express the plan or intention to do something, as we can see in Example 4. (Sense 4).

Example 4:

*a ostatnio **myślę** czy nie zostać krawcową także to się naprawdę może wszystko wydarzyć*

and.lately.ISGthink.whether.not.become.tailor.so.it.(reflexive).really.can. everything happen
[And lately **I've been thinking** whether to become a tailor, so really anything can happen.]

Fifth, speakers of CCC also employed the phrase “I think” to describe the process of deliberating, considering, thinking deeply, and judging reality, as exemplified in Example 5 (Sense 5). In this instance, the subject is the most agentive, as she controls the process and is able to make conclusions with confidence. She is able to induce or stop the process at will.

Example 5:

*ma minę obojętnie ja nie wiem po prostu mam i od razu coś sobie **myślę** na przykład*
has.expression.nevermind.I.do.not.know.just.(I)have.and.instantly.something. **(I)think**.
for.example w związku z tym *aha to on pewnie tak albo taka* in.relation.with.this.aha.
so.he.surely.(is)so.or.(she)is(so).

[Someone is making a face and I – I don't know, this is how I am – and instantly **I am thinking** on the base of it: she is probably like that and he is like that.]

The results of manual annotation of the five senses of the phrase “I think” in the Career Clean Coaching corpus revealed that the most prevalent sense was “Sense 3: to believe in the truth of a proposition”. This sense constitutes 91% of all instances in the corpus (see Table 1).

Table 3: The distribution of the five senses of “I think” in the Career Clean Coaching corpus

Sense	Sense description	English equivalent	Number of instances
Sense 1	to keep in mind	remember	4
Sense 2	to find relationships among things	imagine	3
Sense 3	to believe in the truth of a proposition	think	274
Sense 4	to intend to do something	plan	6
Sense 5	to deliberate, to consider, to think deeply	judge	13
Total instances			300

The results indicate that the most prevalent usage of the phrase “I think” by speakers of CCC is the moderate level of agentivity (Sense 3). Coachees rarely refer to the most agentive and reflective sense 5, where the responsibility for the judgement is high. Conversely, they do not adopt the opposite end of the continuum, as the usage of Sense 1 is not frequent. The reason for this is that the topic under discussion is abstract and complex, namely future career and professional development. Speakers in CCC are not yet prepared to assume full responsibility for their actions, as expressed by the phrase

“I think”, and therefore tend to utilize Sense 3 to mitigate the agentivity of their utterances.

The multiple senses of the Polish expression “I think” reflect the multifunctionality of this utterance. As observed by Danielewiczowa (2002), when “I think” in Polish precedes the proposition (often in the pattern “myślę, że p” – “I think that p”), it serves the function of a hedge, softening the commitment to the claim which follows. Danielewiczowa’s definition aligns with our understanding of Sense 3. This indicates that in the CCC corpus, the expression “I think” is employed primarily in the capacity of a hedge, signifying the coachees’ lack of certainty. This is consistent with the observation of discourse markers indicating hesitation and uncertainty in psychotherapeutic sessions. The presence of discourse markers indicating uncertainty suggests that coaching sessions are similar to psychotherapeutic sessions, as predicted by our application of Ferrara’s model.

4.2. Study 2: The frequency of “I think” in Career Clean Coaching

The objective of this study is to ascertain the frequency of “I think” in the Career Clean Coaching corpus and to compare it with the frequency in other corpora. In order to provide a reference point for the analysis, three corpora of Polish language were utilised. The NKJP, ORIGAMI and PELCRA corpora were used, which represent spoken language.

The NKJP (National Corpus of Polish, Przepiórkowski et al. 2010) is a balanced corpus, comprising both written and spoken language, with a greater prevalence of the former. It contains over 300 million tokens. The ORIGAMI corpus (Karpiński et al. 2008) comprises 9,239 tokens from 40 speakers engaged in task-oriented dialogues (folding paper figures). The spoken data employed in this study was derived from the Polish and English Language Corpora for Research and Applications (PELCRA) (Pęzik 2012). This comprised collections of spoken language, dialogues and polylogues. The topics covered in this corpus are typically informal, including discussions on holiday planning, health, and politics. The participants in the PELCRA corpus provided their consent for the recording of their conversations prior to the commencement of the recording process. However, they were not informed of the precise moment at which the recording would commence.

Table 4: Comparison of frequencies of “I think” in selected corpora of Polish language

Corpus type	Corpus name	Total number of tokens in corpus	Frequency of “I think”	Frequency of “I think” per 10 000 tokens
balanced	NKJP	300 000 000	1000	0,03
spoken	PELCRA	2 126 961	1 119	5,26
	ORIGAMI	9 239	5	5,41
	Career Clean Coaching	104 914	300	28,59

Table 4 presents the frequencies of the verb “I think” in the compared corpora. Given the varying availability and size of spoken and written corpora, the size of our samples also varies. Consequently, the frequency was reported as the number of occurrences per 10,000 words. In the balanced version of the National Corpus of Polish, for the total sample size of 300 million words, we can observe only 1000 instances of the verb “I think”, which gives us a normalized value of 0.03 per 10,000. This indicates that the verb is exceedingly uncommon in this corpus, given that the balanced sample is still predominantly composed of written language. The frequent use of the verb “I think” is a distinctive feature of spontaneous spoken discourse.

In order to ascertain whether the lexical choices of speakers in career coaching sessions are indeed specific to this discourse, or merely a typical feature of spoken language, we conducted a comparative analysis of the CCC corpus with two other types of spoken corpora: The spoken corpora under consideration are PELCRA and ORIGAMI. As illustrated in Table 4, the relative frequency of the verb “I think” in both the compared spoken corpora is comparable, with a value of approximately 5. This suggests that the frequency of use of “I think” remains consistent across both standard, everyday conversations (including dia- and polylogues) and task-oriented dialogues. Nevertheless, the frequency of occurrence of the verb “I think” increases in the context of career coaching sessions, with a rate of 28.69 per 10,000 tokens.

To ascertain whether the observed differences between career coaching sessions and other conversations were statistically significant, we conducted a comparative analysis of the frequencies of the phrase “I think” in the three spoken corpora with the Log Likelihood (LL) value. Comparisons of the CCC corpus (coaching sessions) with the PELCRA (conversational spoken Polish) and ORIGAMI (task-oriented dialogues) revealed statistically significant differences (LL value 478.31, $p < 0.0001$ and LL value 24.75, $p < 0.0001$, respectively). However, the comparison between two other spoken corpora (PELCRA and ORIGAMI) revealed no statistically significant differences in terms of the use of the phrase “I think”.

4.3. Study 3: Speaker differences in Career Clean Coaching

As previous research suggested, variability of most features in spoken corpora is high, and aggregated numbers should be reconfirmed with tests taking between-speaker variability into account (Brezina & Meyerhoff 2014). The data also reflected this high degree of between-speaker variability, as illustrated in Figure 1. The histogram shows the variability in the frequency of the utterance “I think” (normalized per 10,000 tokens) between speakers. Five speakers uttered the phrase between one and five times, with the lowest number being 4.82. Seven speakers uttered the phrase between ten and twenty times, and four speakers uttered it between twenty and thirty times. As the data is highly right skewed, the remaining values are unequally distributed. The highest frequency, found in only one speaker, is 243 instances (normalized per 10,000).

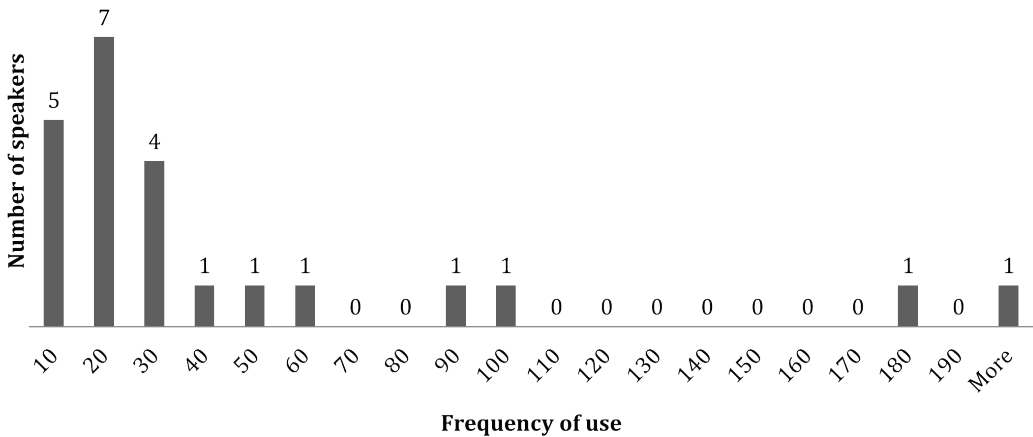


Figure 1: The distribution of frequencies (normalized per 10 000 tokens) of “I think” between speakers in Career Clean Coaching corpus.

To account for the high between-speaker variability, intergroup comparisons were performed using the Mann–Whitney U Test analysis (speakers from CCC versus PELCRA speakers). In order to obtain a reference corpus of a similar size and characteristics to the target (CCC) corpus, a systematic sample was created from PELCRA¹. Currently, PELCRA does not allow for global identification of speakers, thus it is not possible to determine which utterances belong to a particular speaker. However, for each utterance, data about gender, age and education level are provided in the PELCRA results. Consequently, in order to create a corpus with a similar size and gender balance to that of the CCC corpus, it was necessary to create 23 artificial ‘speakers’. The PELCRA sample was divided into 23 ‘speakers’ based on age, gender and education. This resulted in 18 female speakers and 5 male speakers, all aged between 25 and 35 and with a higher education degree. This data was used to conduct a Mann Whitney U Test (see Table 5).

Table 5. A comparison of the frequency of the use of the phrase “I think” between speakers in the CCC and Pelcra corpora is presented in Table 5. The highest frequency of the use of “I think” among female speakers is as high as 243.68 instances per 10,000. However, this speaker was an outlier. As illustrated in Figure 1, the histogram reveals that only two speakers (both female) in the total CCC corpus achieved a value above 100 instances per 10,000 tokens. The second speaker attained a value of 173.61. In the context of everyday conversations (PELCRA corpus), the highest observed value was 19.96 instances per 10,000 (for a female speaker), with the next speaker (male) exhibiting

¹ It was not possible to preserve gender balance in our group due to the lack of male participants volunteering in coaching sessions. This is why for the Mann–Whitney U Test test we created the sample with equally high number of females. In this group the effect of the pragmatic setting (being in the coaching session or conversation) is still statistically significant.

Table 5: Comparison of the frequency of the use of “I think” between speakers in CCC and Pelcra corpus

Speaker ID	Speaker Gender	Career Clean Coaching			PELCRA		Frequency per 10 000
		Words	Sum of “I think”	Frequency per 10 000	Words	Sum of “I think”	
1	Female	5344	11	20.58	3107	3	9.66
2		1055	3	28.44	4157	5	12.03
3		1440	25	173.61	3999	1	2.50
4		3380	6	17.75	5135	2	3.89
5		5246	4	7.62	6538	1	1.53
6		3458	14	40.49	4682	4	8.54
7		11154	9	8.07	4199	2	4.76
8		11202	34	30.35	5668	2	3.53
9		1489	2	13.43	6139	2	3.26
10		6767	67	99.01	3700	4	10.81
11		2726	22	80.70	8285	13	15.69
12		996	1	10.04	3471	1	2.88
13		4016	7	17.43	4168	2	4.80
14		1108	27	243.68	3996	3	7.51
15		5871	30	51.10	4370	1	2.29
16		2289	5	21.84	4709	0	0.00
17		1948	3	15.40	4509	9	19.96
18		6417	12	18.70	4328	2	4.62
19	Male	1381	4	28.96	4185	2	4.78
20		3617	3	8.29	3560	7	19.66
21		1996	2	10.02	3559	2	5.62
22		12454	6	4.82	3899	3	7.69
23		3879	3	7.73	1498	0	0.00
SUM		99233	300	(Av.)30.23	101861	71	(Av.)6.97

a value that was very close to this at 19.66. Figure 2 presents the distribution and outliers between the two groups of speakers. The distribution of the frequency of the utterance “I think” between 23 speakers in the Career Clean Coaching Corpus (left) and the sample of spoken Polish (PELCRA corpus – right) is presented in Figure 2. This figure shows the number of instances of the utterance per 10,000 tokens.

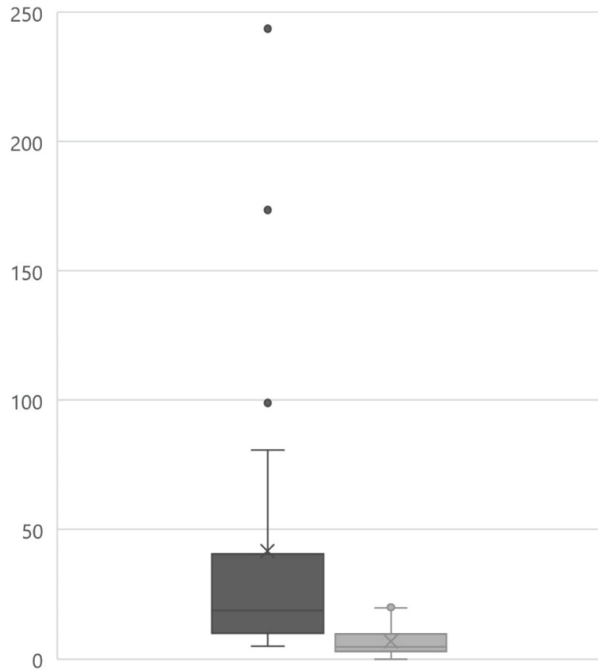


Figure 2: Distribution of the frequency of “I think” between 23 speakers in the Career Clean Coaching Corpus (left) and the sample of spoken Polish (PELCRA corpus – right), normalized per 10,000 tokens

To account for the high between-speaker variability observed in this data, intergroup comparisons were performed using the Mann–Whitney U Test analysis (speakers from CCC versus PELCRA speakers). The U-Mann Whitney test yielded statistically significant differences between two groups of speakers, returning U-values of 61 and 100, respectively, for normalized per 10,000 tokens and actual values. The Z-test was significant at $p < 0.01$, indicating that the two groups of speakers differed with regards to their frequency of use of “I think”, despite between-speaker variability.

5. Discussion

The analysis presented in this paper provides evidence that speakers in specific dialogical situations – career coaching sessions – use the expression “I think” more frequently than similar speakers in task-oriented dialogues or in everyday conversations. Furthermore, the majority of instances of “I think” are used in the function of a hedge, softening the assertive force of the proposition. This allows us to claim that, in addition to individual factors already recognised in the literature, such as gender or age, contextual and pragmatic factors, namely the topic and setting of the conversation, may also influence the use of “I think”. This finding is in accordance with previous studies (Mullan

2010, Ponterotto 2014, Helasvuo 2014, Kärkkäinen 2003, Brezina 2013) – as we found that the frequency of the use of “I think” is highly variable among speakers and may be considered a part of an individual speaker’s linguistic repertoire. Although speakers vary in the frequency of their use of “I think”, in the specific context of coaching sessions, they use this expression at a much higher rate than in other situations. We argue that the conversational setting, along with the specific protocol and abstract and difficult topic, influence speakers’ lexical choices. Additionally, the observed frequency of the use of “I think” as a hedge suggests that this is a distinctive feature of career coaching sessions. During career coaching sessions, speakers tend to grapple with complex and abstract topics such as future career, beliefs and attitudes. We posit that, akin to the metaphor, the frequent use of “I think” in career coaching sessions serves as a linguistic tool for speakers to conceptualize complex subjects, and to signal their uncertainty. This paper analyses one discourse marker of coaching dialogues. Further studies are required to fully understand the nuances of natural, spontaneous dialogue, such as coaching sessions. Individuals discussing abstract concepts, such as their future or career, often feel uncertain and find it challenging to articulate their thoughts. In addition to hedges, other linguistic devices, such as metaphors, should be recognized and analysed as indicators of abstract and complex conversations during coaching sessions.

Conflict of Interest: The authors declare no conflict of interest

Authors’ Contributions: The authors confirm sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Acknowledgements

The paper has been prepared within a research project funded by the *Polish National Science Centre* (decision DEC-2011/03/D/HS6/05993). Authors would like to thank all team members and subjects of the study.

References

- Adler, R.B. & Towne, N. 1996. *Looking out/looking in: Interpersonal communication*. Miasto: Wydawnictwo.
- Baker, P. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baumgarten, N. & House, J. 2010. *I think and I don't know* in English as lingua franca and native English discourse. *Journal of Pragmatics* 42(5). 1184-1200.
- Biber, D. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8(1). 9-37.
- Brezina, V. 2013. *Certainty and uncertainty in spoken language: In search of epistemic sociolect and idiolect*. Frankfurt am Main: Peter Lang.
- Brezina, V. & Meyerhoff, M. 2014. Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics* 19. 1-28. <https://doi:10.1075/ijcl.19.1.01bre>.
- Cameron, L. & Maslen, R. 2010. *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*. London: Equinox Pub.
- Cutting, J. 2000. *Analysing the language of discourse communities*. Amsterdam etc.: Elsevier.
- Danielewiczowa, M. 2002. *Wiedza i niewiedza: Studium polskich czasowników epistemicznych*. Warszawa: Katedra Lingwistyki Formalnej UW.

- Egbert, M. & Yufu, M. & Hirataka, F. 2016. An investigation of how 100 articles in the *Journal of Pragmatics* treat transcripts of English and non-English languages. *Journal of Pragmatics* 94. 98-111.
- Fabiszak, M. 2008. Corpus frequency as a guide to metaphor labelling. In Wąsik, Z. & Komendziński, T. (eds.), *Metaphor and cognition* (Special volume of *Philologica Wratislaviensia: From grammar to discourse*), 149-162. Frankfurt am Main: Peter Lang.
- Ferrara, K.W. 1994. *Therapeutic ways with words*. New York–Oxford: Oxford University Press.
- Freitag, R.M.K. & Cardoso, P.B. & Tejada, J. 2022. Linguistic and paralinguistic constraints on the function of (*eu*) *acho que* as DM in Brazilian Portuguese: A multilevel approach. *Pragmatics & Cognition* 29(2). 324-46. <https://doi.org/10.1075/pc.21024.fre>.
- Górski, R. & Lewandowska-Tomaszczyk, B. (eds.). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Helasvuo, M.L. 2014. Agreement or crystallization: Patterns of 1st and 2nd person subjects and verbs of cognition in Finnish conversational interaction. *Journal of Pragmatics* 63. 63-78.
- Juszczyk, K. 2017. *Mądrość metafory: Komunikacyjny model metafory poznawczej*. Poznań: Coachspace.
- Kaltenböck, G. 2009. Initial *I think*: Main or comment clause. *Discourse and Interaction* 2(1). 49-70.
- Kamasa, V. 2013. Naming “in vitro fertilization”: Critical discourse analysis of the Polish Catholic Church’s official documents. *Procedia – Social and Behavioral Sciences* 95. 154-159.
- Kärkkäinen, E. 2003. *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think*. Amsterdam–Philadelphia: John Benjamins.
- Karpiński, M. & Juszczyk, K. & Jarmołowicz-Nowikow, E. & Malisz, Z. & Szczyszek, M. 2008. Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych. *Investigationes Linguisticae* 16. 83-98.
- Kokorniak, I. & Konat, B. 2012. Agent or experiencer? A search for the subject role in the mental verb *myśleć* ‘think’ in Polish. In Kosecki, K. & Badio, J. (eds.), *Cognitive processes in language*, 77-86. Frankfurt am Main: Peter Lang.
- Konat, B. & Juszczyk, K. 2015. Multimodal communication in career coaching sessions: Lexical and gestural corpus study. *Lodz Studies in Language* 37. 195-212.
- Lakoff, G. 1975. Hedges: A study in meaning criteria and the logic of fuzzy concepts. In Hockney, D. & Harper, W. & Freed, B. (eds.), *Contemporary research in philosophical logic and linguistic semantics* (The University of Western Ontario Series in Philosophy of Science), 221-271. Dordrecht: Springer.
- Lakoff, G. & Johnson, M. 2003. *Metaphors we live by*. 2nd edn. Chicago: University of Chicago Press.
- Lawley, J.D. & Tompkins, P.L. & 2000. *Metaphors in mind: Transformation through symbolic modelling*. 1st edn. London: Developing Company Press.
- Lewandowska-Tomaszczyk, B. 2011. Nowe wyzwania w jakościowej i ilościowej metodologii analizy języka. *Biuletyn Polskiego Towarzystwa Językoznawczego* 67. 51–71.
- McCarthy, M. 1998. *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McMullen, L.M. & Conway, J.B. 1996. Conceptualizing the figurative expressions of psychotherapy clients. In Mio, J.S. & Katz, A.N. (eds.), *Metaphor: Implications and applications*, 59-72. Mahwah, N.J.: Psychology Press.
- Mullan, K., 2010. *Expressing opinions in French and Australian English discourse: A semantic and interactional analysis*. Amsterdam: John Benjamins.
- Oleskeviciene, G.V. & Liebeskind, Ch. 2021. Multiword expressions as discourse markers in Hebrew and Lithuanian. In Bizzoni, Y. & Teich, E. & España-Bonet, C. & van Genabith, J. (eds.), *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, 46-56. Association for Computational Linguistics. (<https://aclanthology.org/2021.motra-1.5>)
- Pawelczyk, J. 2011. *Talk as therapy: Psychotherapy in a linguistic perspective*. Boston–Berlin: De Gruyter Mouton.
- Pęzik, P. 2012. Wyszukiwarka PELCRA dla danych NKJP. In Przepiórkowski, A. & Bańko, M. & Górski, R. & Lewandowska-Tomaszczyk, B. (eds.), *Narodowy Korpus Języka Polskiego*, 253-279. Warszawa: Wydawnictwo Naukowe PWN.
- Pieśkiewicz, B. & Kołodkiewicz, M. 2011. Metafora w coachingu: Zastosowanie metody Clean Coaching w pracy z organizacjami. In Czarkowska, L. (ed.), *Coaching: Katalizator rozwoju organizacji*, 207-223. Warszawa: New Dawn.
- Ponterotto, D. 2014. The risks of uncertainty: Hedging strategies in rape trial discourse. *Language and Dialogue* 4. 93-111. <https://doi.org/10.1075/ld.4.1.06pon>.

- Przepiórkowski, A. & Bańko, M. & Górski, R. & Lewandowska-Tomaszczyk, B. (eds.), *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Sato, S. 2017. On establishing I think as a final particle in interactions: Some comparisons with sentence-final particles in Japanese. *Journal of Pragmatics* 110. 83-98.
- Sloetjes, H. & Wittenburg, P. 2008. Annotation by category-ELAN and ISO DCR. 6th International Conference on Language Resources and Evaluation (LREC 2008).
- Sullivan, W. & Rees, J. 2008. *Clean Language: Revealing metaphors and opening minds*. 1st edn. Carmarthen, Wales: Crown House Publishing.
- Tay, D. 2013. *Metaphor in psychotherapy: A descriptive and prescriptive analysis*. Amsterdam–Philadelphia: John Benjamins Publishing.
- Thorpe, S. & Clifford, J. 2003. *The coaching handbook: An action kit for trainers and managers*. 1st edn. London: Kogan Page.
- Tolvanen, E. 2024. 'I think' in Swedish L1 and L2 group interactions. *Pragmatics* 34(4). 615-641. <https://doi.org/10.1075/prag.22060.tol>.
- Wang, Q. 2020. A corpus-based contrastive analysis of I think in spoken Hong Kong English: Research from the International Corpus of English (ICE). *Australian Journal of Linguistics* 40(3). 319-45. <https://doi.org/10.1080/07268602.2020.1823817>.