

ISO 24617-2 on a cusp of languages

Krzysztof Hwaszcz, Marcin Oleksy, Aleksandra Domogała, and Jan Wiczorek
Wrocław University of Science and Technology

{krzysztof.hwaszcz,marcin.oleksy,aleksandra.domogala,jan.wiczorek}@pwr.edu.pl

Abstract

The article discusses the challenges of cross-linguistic dialogue act annotation, which involves using methods developed for a multilingual framework to annotate conversations in a specific language. The article specifically focuses on the research on dialogue act annotation in Polish based on the ISO standard. To ensure applicability across languages, the standard was designed to be language-independent. The article examines the differences between Polish and English in dialogue act annotation based on selected examples from DiaBiz.Kom corpus, such as the use of honorifics in Polish, the use of inflection to convey meaning in Polish, the tendency to use complex sentence structures in Polish, and the cultural differences that may play a role in the annotation of dialogue acts. The article also discusses the creation of DiaBiz.Kom, a Polish dialogue corpus based on ISO 24617-2:2012¹ standard applied to 1100 transcripts.

1 Introduction: Setting the scene

The process of dialogue act annotation is useful for natural language understanding, speech recognition, and various other applications that require the analysis of spoken language. However, annotating dialogue acts in one language may not be sufficient for processing conversations in another language. In such cases, cross-lingual dialogue act annotation is required, which involves using methods developed for one language to annotate conversations in another language (Bunt et al., 2020; Petukhova et al., 2015).

The ISO 24167-2:2012 standard included native speakers of Belorussian, Dutch, English, French, German, Italian, Japanese, Korean, Romanian, and Swedish. Yet each language has its own specific instruments for expressing communicative functions

¹We also consulted the second edition of the standard: ISO 24617-2:2019.

(and qualifiers). That requires addressing various challenges, such as differences in grammar, syntax, and lexicon, which can affect the accuracy of the annotation process. This paper examines the research on dialogue act annotation in Polish based on ISO standard developed for a multilingual framework (Bunt et al., 2010a). Also, methods used to address these challenges are outlined.

2 Related works

On multiple occasions, when faced with challenges in annotating communication functions, we turned to the literature for inspiration, seeking ideas from existing solutions. For the theoretical background we refer to ISO/DIS 24617-2:2012 (Bunt et al., 2010b, 2012), which is based on particular innovations such as distinguishing between annotations and representations (according to the ISO Linguistic Annotation Framework (LAF, ISO 24612:2009) and sets of dialogue participants, dimensions, communicative functions, functional segments and qualifiers (inventory of DiAML). Both manual and automatic annotation of dialogue segments according to the ISO standard have been tested in practice and described (Keizer et al., 2011; Petukhova et al., 2014; Bunt et al., 2016; Chowdhury et al., 2016; Ngo et al., 2017; Gilmartin et al., 2018). The development of annotation standards for particular corpora can be vividly exemplified by the case of the Switchboard Dialogue Act Corpus (the collection of telephone conversations). The NXT-format Switchboard Corpus was created with additional annotations according to the international standard ISO 64217-2:2012 (Fang et al., 2012). The re-annotation shows the significance of both standard scheme improvement and combining different standards on the same linguistic material.

The DiaBiz.Kom corpus correlates with the DialogBank corpus – current gold annotation stan-

standard. Most dialogues from the DialogBank corpus were taken from other corpora and re-segmented and re-annotated. All annotations were double-checked for inconsistencies, errors and omissions. The data include samples which may be considered illustrative examples for annotations (Bunt et al., 2016). Suggestions and remarks with regard to limitations and extensions of the ISO standard put forth by the authors of the DialogBank have subsequently been implemented in the updated versions of the ISO standard (Bunt et al., 2018)).

3 Polish dialogue corpus DiaBiz.Kom

The DiaBiz.Kom corpus development is an annotation effort performed simultaneously with the DiaBiz corpus creation. The DiaBiz (Pęzik et al., 2022) is a large, multimodal corpus of Polish telephone conversations conducted in varied business settings, comprising 3,766 call center interactions based on 110 business scripts. The recordings were then transcribed and enriched with punctuation. DiaBiz.Kom (Oleksy et al., 2022) was created as an annotation layer based on the ISO 24617-2:2012 standard applied to the 1100 transcripts derived from DiaBiz (10 dialogues for each dialogue script). Every dialogue is annotated by 3 persons: 2 independently working annotators and a super-annotator who resolves all annotation inconsistencies. The authors in the first place focused on communicative function and dimension annotation, then the functional and dependence relations were annotated. Currently, the corpus consists of 1 277 965 tokens (151 520 final annotations for communicative functions). The corpus sample is available under CC BY-NC-ND 4.0 license at: <http://hdl.handle.net/11321/886>.

4 Language differences in dialogue act annotation between Polish and English

Dialogue act annotation is the process of labeling utterances in a dialogue with a specific communicative function or speech act. While Polish and English share common dialogue act categories, such as "inform" and "question," their implementation may bear some noticeable differences between the two languages (Fang et al., 2012; Biały, 2016).

One difference is the use of honorifics in Polish. In Polish the use of formal or informal pronouns and verb forms depends on the relationship between the speakers, their social status, and the context of the conversation. Although the level of

formality in the annotated dialogues between an agent and a client is rather consistent throughout and high, this aspect may still affect the annotation of dialogue acts such as "request", "instruct" or "apology," which may be expressed differently depending on the level of formality required.

Another difference is the use of inflection to convey meaning in Polish. Unlike English, which relies heavily on word order to convey meaning, Polish uses inflection to indicate the grammatical role of a word in a sentence. This can affect the annotation of dialogue acts such as "command," where the inflection of the verb may be more important than the word order.

Additionally, in Polish there is a greater tendency to use complex sentence structures², which can make it more challenging to identify and annotate dialogue acts accurately. In English, there is a preference for simpler sentence structures, which makes it easier to identify and annotate dialogue acts. Moreover, English has a more rigid word order in questions than Polish. In English, the standard word order for questions is to invert the subject and auxiliary verb and add a question word or particle at the beginning or end of the sentence (1). Polish, on the other hand, allows for more flexibility in word order in questions. While the standard Polish word order for questions is similar to that of English (2a), Polish also allows for alternative word orders depending on the emphasis or focus of the question: inverted word order without a question particle (2b) and inverted word order with the verb and object reversed (2c).

(1) Do you like pizza?

(2)
a. Czy lubisz pizzę?
b. Lubisz pizzę?
c. Pizzę lubisz?

Finally, cultural differences may also play a role in the annotation of dialogue acts in Polish and English. For example, in Polish culture the intended meaning may not be explicitly stated, but rather implied through context and cultural norms. This can make it more challenging to identify and annotate dialogue acts accurately in Polish. Also, similar communication behavior may take different forms. An example is the beginning of a conversation, in

²The assertion is based on the study conducted by Ostalak, who examined grammatical structures in sentences within formal topics (Ostalak, 2019). The author established that complex sentences in Polish constituted 26,23%, while these in English – 18,97%

which the interlocutors establish their positions in the conversation, for example, one of them shows willingness. In English, it is typically an interaction:

“Can I help you?” [Offer] ↔ “Yes.” [acceptOffer]

In contrast, in Polish it is usually a less elaborate, unidirectional structure. The expression of willingness to help is only in the form of a question. Essentially, it is an encouragement to the caller to state his or her problem right away. For this reason, we decided to label such segments as Interaction-Structuring.

5 Morphological richness in the language and its influence on annotations

Dialogue act annotation in morphologically rich languages, such as Polish, can be more challenging than in morphologically poor languages, such as English. This is because morphologically rich languages have a greater number of inflections and grammatical markers that can affect the interpretation of utterances. In morphologically rich languages the inflection of a word can change its meaning or grammatical function, which can have implications for the annotation of dialogue acts. For example, the use of a different verb form can indicate whether a request is polite or imperative. The use of case markers can also indicate the role of a noun in a sentence, which can affect the annotation of dialogue acts such as "offer" or "request."

Morphologically poor languages, such as English, have fewer inflections and grammatical markers, which can make it easier to identify and annotate dialogue acts. English relies more on word order and lexical cues to convey meaning, which makes it easier to identify the main clauses and subordinate clauses in a sentence.

However, morphologically poor languages like English also have their own challenges in dialogue act annotation. For example, English often uses indirect speech acts, where the intended meaning is not explicitly stated, but rather implied through context and cultural norms. This can make it more challenging to identify and annotate dialogue acts accurately in English. However, it is still possible to achieve accurate and reliable results with careful annotation guidelines and a thorough understanding of the language's grammar and syntax.

6 Multipolysemous words: Selected examples

Another issue that adds up to the challenges of dialogue act annotation in Polish is the notion of polysemous words. Such words have multiple meanings that can lead to ambiguity in their interpretation (Gruszczyńska et al., 2019; Lewandowska-Tomaszczyk and Thelen, 2013). When annotating dialogue acts, it is important to disambiguate the meaning of polysemous words to ensure that the intended dialogue act is accurately labeled (Silvano et al., 2022).

While annotating problematic examples, we were more likely to interpret words with identical orthographic form that have distinct meaning as different lexical units rather than as polysemous words. Thus, we were closer to a structural perspective (Apresjan, 1974; Bogusławski, 1976), than, for instance, a cognitive one (Lakoff, 2008). One challenge is identifying the context in which the polysemous word is used. Context plays a crucial role in disambiguating the meaning of a word (Mohammed, 2009; Schmidt, 2008). Therefore, it is important to consider the surrounding words and the larger context of the dialogue when annotating dialogue acts.

Annotators may also need to rely on their own personal (and subjective) knowledge and experience to disambiguate the meaning of polysemous words. This can be especially challenging when annotating dialogues that cover a wide range of topics.

Another challenge is to distinguish between the various meanings of a polysemous word. Some polysemous words may have meanings that are closely related, making it difficult to differentiate between them.

To mitigate these challenges, it is important to provide annotators with clear guidelines and instructions that specify how to disambiguate polysemous words. These guidelines should also include examples and explanations of how to interpret and annotate these words in different contexts. We have also developed substitution tests for selected cases. Additionally, it was helpful to have multiple iterations as well as annotator reviews and discuss the annotations to ensure consistency and accuracy in the labeling of dialogue acts.

Let us consider below a number of examples of polysemous words along with the approach we have adopted on the basis of annotator domain-

specific knowledge and experience, tests as well as additional contextual information

6.1 “Proszę”

The most common translation of the Polish word “proszę” into English is “please”. However, depending on the context and usage, it may also be translated as “thank you”, “you’re welcome”, “excuse me” or “here you are”. The range of possible translations inevitably triggers the number of functions it may take when annotating dialogue acts.

“Proszę” in Polish and “please” or “you’re welcome” in English can be categorized as markers of politeness or as a response to a polite request or gratitude. In dialogue act annotation, the specific meaning of “proszę” would depend on the context and the speaker’s intention. For instance, “proszę” in a phrase could be annotated as a Request (e.g., “Proszę mi pomóc”, Eng. “Please help me.”), or as an Accept Thanking in the dialogue turns like - “Dziękuję. - “Proszę bardzo.” (Eng. - “Thank you” - “You’re welcome”). In English, these two functions – namely Request and Accept Thanking would be split between two separate lexical items, the former function being reserved for “please”, whilst the latter for “you’re welcome”.

All in all, we have distinguished five different dialogue functions (Accept Thanking - 19 cases, Accept Offer - 7, Contact Indication - 30, Contact Check - 1, Auto Negative - 1) for “proszę” illustrated in (3)-(7) below.

(3)

Agent: Dobrze. **Proszę**. [acceptThanking: SOM] “Agent: Alright. **You’re welcome.**”

Klient: (...) Dobrze. **Dziękuję bardzo**. [thanking: SOM] “Client: (...) Good. Thank you very much.”

(4)

Agent: Jasne. Tak, **mogę, mogę złożyć taką dyspozycję wyłączenia**. [Offer: Task] “Agent: Sure. Yes, I can, I can make such an exclusion order.”

Klient: **Proszę**. [acceptOffer: Task] “Client: **Please.**”

(5)

Klient: **Proszę**. [contactIndication: Contact Management/ opening: Discourse Structuring] “Client: **Hello.**”

Agent: (yy) Dzień dobry. Czy ja dodzwoniłam się do pani...”Agent: (yy) Good morning. Have I reached Mrs...

(6)

Klient: (yy) siedemdziesiąt trzy, zero, osiem, zero, dwa. “Client: (yy) seventy-three, zero, eight, zero, two.”

Agent: **Proszę... proszę...** [contactCheck: Contact Management] “Agent: **Go on... go on... / Yes... yes...**”

Klient: Zero, osiem... “Client: Zero, eight...

(7)

Klient: Tak, tak. “Client: Yes, yes.”

Agent: To tak. “Agent: That’s right.”

Klient: **Proszę?** [autoNegative: Auto-Feedback] “Client: **Excuse me?**”

6.2 “Dobrze”

The word “dobrze” in Polish does not have its one English counterpart – the meaning lies on the verge of “okay” and “alright” (as expressions of agreement or acceptance), which makes it even more difficult to compare the two languages. In Polish “dobrze” can be used to indicate agreement, approval, or satisfaction, but it can also be used to indicate understanding or comprehension. In English, “okay” or “alright” are generally used to indicate agreement or acceptance, but – unlike in Polish – they may also be used to indicate indifference or lack of enthusiasm. The decision of the speaker to use “dobrze” in Polish can also depend on the social and cultural context of the conversation. For example, “dobrze” can be used to indicate politeness or deference to a speaker who is perceived to be of a higher social status. In English, “okay” or “alright” are generally used regardless of social context, but can be used to express politeness or informality depending on the situation. “Dobrze” in Polish can also imply a sense of satisfaction or contentment with the situation or outcome. It can also suggest a positive evaluation or endorsement of something. In English, “okay” or “alright” generally do not carry the same level of positive evaluation or endorsement. We have distinguished three different dialogue functions (Auto Positive - 3111 cases, Accept Request/Offer/Suggest - 400, Contact Indication - 12) for “dobrze” illustrated in (8)-(12) below.

(8)

Klient: Tak jest, dokładnie. Wróblewskiego szesnaście jest. “Client: Yes, exactly. It is Wróblewskiego sixteen.”

Agent: **Dobrze**. [autoPositive: Auto-Feedback] “Agent: **Good.**”

(9)

Agent: **Ale proszę się jeszcze tam skontaktować**, [Suggest: Task] (...)”Agent: **But please still get in touch there, (...)**

Klient: **Dobrze**. [acceptSuggest: Task] Dziękuję. “Client: **Okay.** Thank you.”

(10)

Agent: Dobrze, **proszę o chwilę cierpliwości** [Request: Time Management] “Agent: Well, **please be patient for a moment**”

Klient: **Dobrze**. [acceptRequest, wymiar: Time Management] “Client: **Okay.**”

(11)

Agent: (...)proponuję, abyśmy tutaj (...) wspólnie, w trakcie trwania połączenia, wystawili, wystawili reklamację do tej faktury, dobrze? [Offer: Task] "Agent: (...) I propose that we here (...) together, during the call, issue a claim to this invoice, alright?"

Klient: **Dobrze.** [acceptOffer: Task] "Client: **Alright.**"

(12)

Agent: Czterdzieści trzy. Tak? "Agent: Forty-three. Yes?"

Klient: Tak. "Client: Yes."

Agent: **Dobrze.** [contactIndication: Contact Management] "Agent: **Go on.**"

6.3 "Tak"

The third word we wish to consider is the word "tak" in Polish, which may be roughly translated as "yes" in English.

In Polish "tak" can be used in a variety of situations, including to answer yes or to acknowledge understanding. It can also be used as a discourse marker to indicate agreement, to signal a willingness to continue the conversation, or to show politeness. In English, "yes" is generally used to answer a question or to indicate agreement or affirmation.

The use of "tak" in Polish can also depend on the social and cultural context of the conversation. For example, "tak" can be used to indicate politeness or deference to a speaker who is perceived to be of a higher social status. In English, "yes" is generally used regardless of social context, but can be used to express politeness or informality depending on the situation. The word "tak" in Polish can also imply a level of certainty or emphasis in agreement or affirmation. It can also suggest that the speaker is more committed to their agreement or affirmation than "yes" in English. In contrast, "yes" in English is generally more neutral in terms of emphasis or certainty. We have distinguished four different dialogue functions (Confirm - 739 cases, Contact Indication - 1311, Auto/Allo Positive - 257, Agreement - 326) for "tak" illustrated in (13)-(16) below.

(13)

Klient: **Autobus 121 odjeżdża z rogu Podleśnej w kierunku Wrzeciona, prawda?** [checkQuestion: Task] "Client: **Bus 121 leaves from the corner of Podleśna towards Wrzecion, correct?**"

Agent: **Tak.** [Confirm: Task] "Agent: **Yes.**"

(14)

Klient: A czy... "Client: And is..."

Agent: **Tak?** [contactIndication: Contact Management] "Agent: **Yes?**"

Klient: Czy to wtedy (yy) przyjdzie ktoś osobiście... "Client: Is it then (yy) that someone is going to come in person..."

(15)

Agent: Wystawiła pani trójkąt i co najmniej sto metrów przed pojazdem pani postawiła. "Agent: You pulled out a warning triangle and put it out at least a hundred metres in front of your vehicle."

Klient: **Tak.** [alloPositive: Allo-Feedback] "Client: **Yes.**"

(16)

Klient: (yy) Aż tyle mam możliwości. "Client: (yy) So many possibilities."

Agent: **Tak.** [Agreement: Task] "Agent: [Yes.]"

7 Conclusion / General discussion

Dialogue act annotation involves assigning a specific communicative function or speech act to each utterance in a conversation. The process of annotating dialogue acts can be affected by differences in language between Polish and English. In Polish honorifics and inflection are commonly used to convey meaning, which can make it difficult to accurately identify and annotate dialogue acts. The complexity of Polish sentence structures can also present a challenge. Morphologically rich languages, like Polish, have more inflections and grammatical markers that can complicate the annotation process. Additionally, polysemous words can create confusion and ambiguity when trying to distinguish between multiple meanings. To address these challenges, clear guidelines and instructions should be provided to annotators, and multiple rounds of reviews and revisions should be performed to ensure accuracy and consistency. Context and cultural norms can also be helpful in disambiguating the meaning of polysemous words.

The ISO standard serves as a suitable framework for annotating dialogues in different languages. The existing categories provided a means to address cases where language differences emerged, while considering the contextual factors played a crucial role in reaching final decisions. To ensure the adoption of specific solutions, it is important to maintain a consistent approach to dimension recognition. Given the variations across languages, the ISO standard should have a well-defined theoretical foundation, as English examples may not always be sufficient.

To enable effective utilization of the model, guidelines are necessary, empowering annotators to conduct a comprehensive analysis that incorporates both conceptual frameworks and specific textual structures. This entails providing clear and practical definitions of annotated categories, establishing a solid theoretical basis, as well as discussing illustrative examples.

References

- Ju D Apresjan. 1974. *Leksičeskaja semantika: Sinonim-ičeskije sredstva jazyka*. Nauka.
- Andrzej Bogusławski. 1976. O zasadach rejestracji jednostek języka. *Poradnik językowy*, 8(342):356–364.
- Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010a. [Towards an ISO Standard for Dialogue Act Annotation](#). In *Seventh conference on International Language Resources and Evaluation (LREC'10)*, La Valette, Malta.
- Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010b. [Towards an ISO standard for dialogue act annotation](#). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. [ISO 24617-2: A semantically-based standard for dialogue annotation](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437, Istanbul, Turkey. European Language Resources Association (ELRA).
- Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Chengyu Fang, Simon Keizer, and Laurent Prévot. 2020. The iso standard for dialogue act annotation, second edition. In *International Conference on Language Resources and Evaluation*.
- Harry Bunt, Volha Petukhova, Andrei Malchanau, Kars Wijnhoven, and Alex Fang. 2016. [The DialogBank](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3151–3158, Portorož, Slovenia. European Language Resources Association (ELRA).
- Harry Bunt, James Pustejovsky, and Kiyong Lee. 2018. [Towards an ISO standard for the annotation of quantification](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Shammur Absar Chowdhury, Evgeny Stepanov, and Giuseppe Riccardi. 2016. [Transfer of corpus-specific dialogue act annotation to ISO standard: Is it worth it?](#) In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 132–135, Portorož, Slovenia. European Language Resources Association (ELRA).
- Chengyu Fang, Jing Cao, Harry Bunt, and Xiaoyue Liu. 2012. The annotation of the switchboard corpus with the new iso standard for dialogue act analysis. In *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation*. Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation ; Conference date: 03-10-2012 Through 05-10-2012.
- Emer Gilmartin, Christian Saam, Brendan Spillane, Maria O'Reilly, Ketong Su, Arturo Calvo, Loredana Cerrato, Killian Levacher, Nick Campbell, and Vincent Wade. 2018. [The ADELE corpus of dyadic social text conversations:dialogue act annotation with ISO 24617-2](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Ewa Gruszczyńska, Małgorzata Guławska-Gawkowska, and Anna Szczęsny. 2019. *Translatoryczne i dyskursywne oblicza komunikacji*. Instytutu Lingwistyki Stosowanej WLS UW.
- Simon Keizer, Harry Bunt, and Volha Petukhova. 2011. [Multidimensional Dialogue Management](#), pages 57–86. Springer Berlin Heidelberg, Berlin, Heidelberg.
- George Lakoff. 2008. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.
- Barbara Lewandowska-Tomaszczyk and Marcel Thelen. 2013. Translation and meaning part 10.
- Essa T. Mohammed. 2009. Polysemy as a lexical problem in translation.
- Thi-Lan Ngo, Son-Bao Pham, Khac-Linh Pham, Xuan-Hieu Phan, and Minh-Son Cao. 2017. [Dialogue act segmentation for vietnamese human-human conversational texts](#). In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, pages 203–208.
- Marcin Oleksy, Jan Wieczorek, Dorota Drużyłowska, Julia Klyus, Aleksandra Domogała, Krzysztof Hwaszcz, Hanna Kędzierska, Daria Mikoś, and Anita Wróż. 2022. Diabiz.com - towards a polish dialogue act corpus based on iso 24617-2 standard. In *International Conference on Computational Linguistics*.
- Mateusz Arkadiusz Ostalak. 2019. *A comparative analysis of grammatical structures and vocabulary in Polish and English Facebook chats*. Katowice.
- Volha Petukhova, Harry Bunt, Andrei Malchanau, and Ramkumar Aruchamy. 2015. Experimenting with grounding strategies in dialogue. *SEMDIAL 2015 goDIAL*, page 198.
- Volha Petukhova, Martin Gropp, Dietrich Klakow, Gregor Eigner, Mario Topf, Stefan Srb, Petr Motlicek, Blaise Potard, John Dines, Olivier Deroo, Ronny Egeler, Uwe Meinz, Steffen Liersch, and Anna

Schmidt. 2014. The DBOX corpus collection of spoken human-human and human-machine dialogues. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 252–258, Reykjavik, Iceland. European Language Resources Association (ELRA).

Piotr Pęzik, Gosia Krawentek, Sylwia Karasińska, Paweł Wilk, Paulina Rybińska, Anna Cichosz, Angelika Peljak-Lapinska, Mikołaj Deckert, and Michał Adamczyk. 2022. Diabiz – an annotated corpus of polish call center dialogs. In *International Conference on Language Resources and Evaluation*.

Goran Schmidt. 2008. Polysemy in translation - selecting the right sense.

Purificação Silvano, Mariana Damova, Giedrė Valūnaitė Olekeviienė, Chaya Liebeskind, C. Chiarcos, Dimitar Trajanov, Ciprian-Octavian Truică, Elena Simona Apostol, and Anna Bączkowska. 2022. Iso-based annotated multilingual parallel corpus for discourse markers. In *International Conference on Language Resources and Evaluation*.