

BETWEEN ADVERBS AND PARTICLES: A CORPUS STUDY OF SELECTED INTENSIFIERS

JANA ŠINDLEROVÁ¹ – BARBORA ŠTĚPÁNKOVÁ²

¹ Institute of Theoretical and Computational Linguistics, Faculty of Arts, Charles University, Prague, Czech Republic

² Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

ŠINDLEROVÁ, Jana – ŠTĚPÁNKOVÁ, Barbora: Between adverbs and particles: A corpus study of selected intensifiers. *Journal of Linguistics*, 2021, Vol. 72, No 2, pp. 444 – 453.

Abstract: In this paper, we present a preliminary study of three intensifiers (*absolutně, naprosto, úplně*) based on data from three different corpora, a written corpus SYN2020, a web corpus ONLINE-ARCHIVE, and a spoken corpus ORTOFON 1. Providing a parallel annotation of a random sample of each intensifier, we focus on their functions and meanings in context. We analyse their properties in order to define those features which are relevant to their word class assignment, and to prepare grounds for the future disambiguation tasks.

Keywords: particles, adverbs, intensifiers, corpus, Czech

1 MOTIVATION

“Intensifiers”¹, i.e., words strengthening the meaning of the words in their scope, appear problematic with respect to the word class affiliation of the individual candidate words, i.e., their affiliation either to the class of particles, or to the class of adverbs. So far, the individual studies (including relevant chapters in grammars) concerned with words like *trochu, velice, úplně* etc. evaluate them differently. This is particularly interesting, because adverbs and particles (as they are defined) should differ both in their syntactic function and in their semantic interpretation. While in Czech linguistic tradition adverbs always work as syntactic constituents, particles do not, they are believed to operate in higher linguistic layers and to acquire pragmatic functions. While adverbs are considered semantic words, having a full meaning, particles are described as synsemantic ones, having a weakened or modal meaning.²

¹ We use the term intensifier in this study to avoid referring to all the investigated words in terms of word class categories.

² The term *particles* refers here to the category of words expressing the pragmatic dimension of the utterance, as it is traditionally defined in Central European linguistics. The term thus does not apply to words in the function of grammatical operators, e.g., the reflexive element *se* in *dívá se* or *to in to be*.

The task of identifying particle uses of intensifier words, and distinguishing them from adverbial uses is needed e.g., for a consistent morphological tagging of linguistic corpora. For example, the disambiguation processes in the current SYN2020 corpus [1] (using the MorfFlex dictionary [2]) almost do not involve particle/adverb rules.

Since the available theoretical studies do not offer a satisfying and thorough argumentation on how to treat intensifying words, a corpus study is needed to describe behaviour of such words and offer corpus-driven criteria to support their word-class categorization.

In this paper, we present a preliminary corpus study aimed at three representative intensifying words: *absolutně*, *naprosto*, and *úplně* ('absolutely, completely, totally'). In their lexicographic treatment, they are often presented as near synonyms, expressing similar meanings and appearing in similar contexts, cf. SSČ [3]. We investigate their function, their meaning, and their context in three different corpora, and based on a pilot annotation, we point out features leading to difficulties in the task of word class disambiguation.

2 PARTICLES VERSUS ADVERBS: THE THEORY

There is a considerable lack of criteria to delimit the category of particles as a unified and compact system. The existing criteria are largely negative in nature. Particles are primarily a) inflective, b) synsemantic, and c) they do not function as a clause constituent (see e.g., [4, p. 90]). A further delimitation of particles in contrast to other synsemantic word classes works on the basis of elimination: they do not assign case, they do not conjoin words or clauses. As category-unifying features, mostly the following are presented: ability to modify clauses, ability to link the proposition with the context, and expressing the relation of the speaker to the communication situation (cf. e.g., MČ2 [5, p. 228] or VAGSČ [6, p. 91]). There are attempts to define semantically compact subclasses within the category, nevertheless, the individual authors differ in the number and extent of the subclasses distinguished.

One of the subclasses sometimes identified within the class of particles are the intensifiers, i.e., words like *velmi*, *zcela*, *úplně*. Before establishment of particles as a separate category, they were generally considered adverbs. Moreover, some of them also still hold a separate adverbial meaning (e.g., they can be used as obligatory adverbial complementations of verbs, etc.). Therefore, they are sometimes treated as adverbs of measure in literature.

MČ2 [5] treats intensifiers under the label of measure (or intensification) adverbs, which is considered a subclass of manner adverbs. The measure adverbs are given as a list, without detailed characterization or contextual exemplification (p. 190). Additionally, intensifiers are treated in the chapter of particles as well, this time as a subclass of "measure evaluating particles". Again, the potential members

of the class are listed only and the list overlaps with measure adverbs. A similar approach is offered in PMČ [7].

It is quite clear that even when looking into a single grammar of the Czech language, it may not be obvious where the borderline between two distinct uses of the same intensifier is. The underlying cause of the ambiguous treatment of the so-called measure adverbs and intensifying particles is the problem of defining a difference between intensification and emphasis, a subject broached in the Czech linguistic discourse as early as starting with Mathesius (1947; [8]), who acknowledges their overlapping character. Nekula (PMČ [7, p. 360]) mentions that in some cases, the word class affiliation can be influenced by the relative position of the intensifier to the affected word. Vondráček in VAGSČ [6, p. 103] summarizes his findings from a separate study on the topic [9] in the following way: the criterion for distinction of the two word classes lies in the dominance of either the function of specifying measure, or the generally modifying function, together with the ability of the intensifier to work as a syntactic complementation. A similarly general solution is offered by Šimková (2002; [10]).

Looking at the lexicographic approaches to word class assignment to intensifiers, the older monolingual dictionaries SSČ [3] and SSJČ [11] usually assign a single label, either a particle, or an adverb. An attempt to differentiate between the two word classes through separate entries can be seen in the newly prepared ASSČ [12] on the example of *absolutně*, which is presented as an adverb, and also a particle.

3 DATA ANNOTATION

3.1 Annotation process

As a data resource, we have chosen three different corpora representing three different types of text: a representative written corpus SYN2020 [1], a web corpus ONLINE_ARCHIVE [13] and a representative corpus of spoken language ORTOFON [14]. For each of the selected intensifiers (*absolutně*, *naprosto*, *úplně*), we have obtained 50 random concordances.³

Each of the intensifiers was annotated for the following features: the word class of the word in the scope of the intensifier; the position of the intensifier relative to the word in its scope; the function of the affected word within the clause; the position of the intensifier within the clause; the intensifier's assumed word class.

Most concordances included an intensifier modifying an adjective or a verb, consequently, the most frequent syntactic functions of the word in the scope were verbonominal predicates, verbal predicates, and attributes. Majority of cases were intensifiers in the anteposition to the affected words, most prominently in

³ The intensifier *absolutně* gives only 45 concordances in the ORTOFON corpus overall.

a penultimate position in the clause. The spoken corpus showed a notable number of intensifier postpositions, related to a higher number of clause-final positions.

The annotation of syntactic features also served the purpose of making sure that both annotators interpreted the meaning of the sentence in the same way, which is especially important in the case of web and spoken corpora. The results showed that clearly different interpretations were indeed rare.

3.2 Inter-annotator agreement

We calculated the inter-annotator agreement simply as a proportion of the number of cases where the annotators agreed on the label assigned to the number of cases where they differed in the assigned labels. For this pilot study, we were interested mainly in the overall certainty of the annotators to assign the labels and whether the points of disagreement share some common semantic or syntactic features.

Points of disagreement in the word class affiliation of the affected words were infrequent and concerned mainly delimitation of the scope in the case of verbonominal predicates or complex predicates. This often resulted in disagreement in the syntactic function of the affected complementation. None of the differences in judgement of morphosyntactic properties of the affected word seems to have had a direct impact on the word class assignment to the intensifier itself.

Our main interest lied in the agreement on the word class categorization of the intensifier itself. We hypothesized that in view of the fact that the current linguistic theories do not offer a satisfying account of what constitutes a particle, the inter-annotator agreement in this task would be rather low. This was confirmed by the data. In the SYN2020 [1] and ONLINE [13] data, the agreement was lower than 75 %; the annotators chose different labels in more than a quarter of the occurrences.

Intensifier	<i>absolutně</i>			<i>úplně</i>			<i>naprosto</i>		
	D	T	DA	D	T	DA	D	T	DA
SYN2020	10	32	6	19	11	20	31	3	16
ONLINE	7	34	9	22	13	15	25	6	17
ORTOFON	8	31	6	10	28	10	23	19	8

Tab. 1. Annotator agreement⁴

Only a few, syntactically restricted types of contexts allow a definite agreement on the word class assignment. E.g., if the intensifier modifies a syntactic noun

⁴ We use D for adverbs, T for particles, and DA for disagreement. The numbers in the tables in some cases do not add up to the total number of concordances evaluated. We did not include into the overall counts cases when the word class assignment was not possible due to the utterance being too fragmented.

(nouns, totalizing or indefinite pronouns, rarely an infinitive) in the function of a syntactic subject or object, the intensifier is annotated as T.

The word class affiliation of adjectives (whether in the syntactic function of an attribute, or as the nominal part of the verbonominal predicate) is rather unconditioned by the morphological and syntactic interpretation of the word in context. The decision is thus probably driven by semantic factors. This is rather bad news for the efforts to formulate contextual disambiguation rules for automatic processes.

The fact that the same intensifiers might be assigned both the adverb label and the particle label in similar contexts, similar clause positions, and similar syntactic constructions indicates that the process of word class recategorization (from adverb to particle) has not been yet completed. Nevertheless, the annotation and the following analysis show certain semantic tendencies.

A significantly higher agreement in the *absolutně* data suggests that the process of particularization of the expression is already advanced.

The agreement is considerably better in the spoken corpus data. Also, in spoken data, the particle label (T) is more frequently assigned considering the otherwise ambivalent intensifiers *naprosto* and *úplně*. This is probably caused by the fact that spoken data tend to include more clear attitude markers than the written ones.

3.3 Selected points of disagreement in the intensifier word class

The certainty of the annotators varied regarding each individual intensifier. Whereas *absolutně* achieves the highest agreement (85 %), *naprosto* and *úplně* reach around 70 %. While the highest agreement was observed in the spoken corpus, it seems that the disambiguation issue may be supported by the prosodic properties of the utterance, or by other features of a spoken text (simple structures etc.).

Typically, the disagreement arises in combination with evaluative expressions, with qualitative adjectives and with positive verb forms.

By evaluative expressions, we mean primarily qualitative qualification adjectives or adverbs expressing subjective evaluation (1), cf. Karlík in NESČ [15].

- (1) *Školním divadlem jsem byla naprosto nadšená.* ‘I was totally excited about the school theatre performance.’

Here, the annotation basically confirms Vodráček’s opinion [9] that distinguishing adverbial from particle meaning is often dependent on “the semantic interpretation of the modified or specified expression”, i.e., whether the intensifier in context expresses measure circumstance or speaker’s attitude. Nevertheless, our annotation data suggest that the decision is rather subjective and based probably on whether the annotator perceives the expression in the given context as scalable or as a representation of an upper limit which is only emphasized, see an example of a disagreement in (2).

- (2) *to je úplně jednoduchý* ‘That’s totally simple.’ (ORTOFON)

Whereas *absolutně* modifying a negative verb is assigned a T label almost uniformly, positive verb forms trigger both measure interpretation (intensification) and emphasizing interpretation (3).

- (3) *absolutně mě dokáže odradit chlap, který je blbý* ‘I feel absolutely appalled by a guy who is dull.’ (ONLINE)

Naprosto and *úplně* seldom affect negated verbs, therefore, the disagreement concerns mainly combinations with positive verb forms.

4 INTENSIFIERS IN CONTEXT

4.1 Absolutně

The word *absolutně* appears to be strongly tied to negative contexts. By negative contexts, we mean collocations with explicit morphological negation (4), as well as collocations with words and phrases with negative meaning but no morphological marking (5).

- (4) *...absolutně jsem netušil, jak se bude můj život dál vyvíjet.* ‘I had absolutely no idea how my life would go on.’ (SYN2020)
- (5) [Je] *Absolutně vyloučeno, abych zabloudil.* ‘[It’s] Absolutely out of question for me to get lost.’ (SYN2020)

The tendency to be bound to negative context is the strongest in ORTOFON (40 out of 45), the corpus of informal spoken Czech, (42 out of 50) while in SYN2020, a corpus of written texts, the number of positive collocations rises and negative contexts reach only 31 out of 50. This fact constitutes one of the major function differences between *absolutně* on one hand and *úplně* or *naprosto* on the other hand, making it close to *vůbec*, an accepted Czech “negative polarity item”.⁵

The positive contexts include the following types: the intensification of evaluative adjectives or adverbs (6), the intensification of words expressing sameness or different character (*stejný, jinak*) (7), or modification of an objective quality (8).

- (6) *Auto, které vidíte na obrázku před sebou se jmenuje Interceptor S a je absolutně skvostné.* ‘The car you see in the picture in front of you is called Interceptor S and is absolutely brilliant.’ (SYN2020)

⁵ *Absolutně*, though, cannot be considered an NPI, since it fits perfectly into positive contexts: *S názvem článku absolutně nesouhlasím.* ‘I absolutely do not agree with the article title.’ *S názvem článku absolutně souhlasím.* ‘I absolutely agree with the article title.’

- (7) *Půjde o absolutně jiný film.* ‘It will be an absolutely different film.’ (SYN2020)
- (8) *Stal jsem se na ní absolutně závislej.* ‘I became absolutely dependent on her.’ (SYN2020)

Rarely, the word *absolutně* appears in the function of a clear manner adverbial (9).

- (9) *Pokud řada nekonverguje absolutně, může její součet být v rozporu s naším očekáváním.* ‘If a series does not converge absolutely, its sum may be contrary to our expectation.’ (SYN2020)

As for the position of the word within a sentence, while in SYN2020 and ONLINE, *absolutně* stands almost exclusively in front of the word in its scope, sometimes in the middle position (usually with copular predicates), the ORTOFON data show 7 cases of a postposition (10). In case the sentential stress lies on the verb in the scope, the word is then easily interpreted as a particle.

- (10) *Jo to já ti rozumím absolutně.* ‘Yeah, I absolutely understand you.’ (ORTOFON)

4.2 Naprosto

In contrast to *absolutně*, *naprosto* does not appear significantly in negative contexts, though a negative context is not excluded (11).

- (11) *Zatím bohužel rozšiřuje řady kočiček, kterým útulek naprosto nevyhovuje.* ‘Unfortunately, it is increasing the number of cats, for which a shelter is not convenient at all.’ (ONLINE)

Naprosto often modifies positive evaluative words (12), but it also appears with words describing sameness or different character (13).

- (12) *Vše se nese v naprosto pohodové atmosféře.* ‘Everything is carried on in a totally relaxed atmosphere.’ (ONLINE)
- (13) *Barevnost nechali čistě na nás, měli jsme však vybrat naprosto odlišné odstíny, než byly v původní ložnici.* ‘They left the color purely on us, but we had to choose totally different shades than those of the former bedroom.’ (SYN2020)

Also, in comparison to the other two intensifiers discussed, *naprosto* is notably more often used even with non-evaluative words (i.e., words expressing some quality primarily without judging whether the quality is positive or negative) (14).

- (14) *Vše tkví ve výchově a vzdělání a v těchto dvou oblastech si mohou být ženy s muži naprosto rovný.* ‘It’s all about upbringing and education, and women can be totally equal to men in these two areas.’ (SYN2020)

4.3 Úplně

Úplně, in contrast to the previous candidates, almost does not appear with negative verbs in our data. This is probably connected to the fact that when combined with a negated verb, its meaning is shifted. While *naprosto* expresses an upper limit, a total completion of the verb meaning with positive verbs (15a), and a lower limit, a total incompleteness of the verb meaning with negative verb (15b), *úplně* expresses an upper limit, a total completion of the verb meaning with positive verbs (15a), but with negative verbs in some syntactic contexts, its meaning is rather reaching a low level of completion of the verb meaning (15c).

- (15a) *Naprosto/Úplně s tebou souhlasím.* ‘I strongly agree with you.’

- (15b) *Naprosto s tebou nesouhlasím.* ‘I strongly disagree with you.’

- (15c) *Úplně s tebou nesouhlasím.* ‘I slightly disagree with you.’

Úplně in this context weakens the negative meaning, rather than intensifying it.

In positive contexts, *úplně* more or less shares the usual collocations of *naprosto* and *absolutně*. It combines with the sameness or difference expressions more often than *naprosto* (16).

- (16) *tam je úplně jiná mentalita ještě* ‘there is a completely different mentality yet’ (ORTOFON)

It combines with positive totalizing pronouns (*všechno, každý*), but negative pronouns are extremely rare in its scope, mostly they are perceived as rather incompatible, because the natural opposite to *úplně* in negative contexts is *vůbec* (17).

- (17) *Neměl jsem vůbec/??úplně žádné peníze.* ‘I had no money at all.’

We can find examples of *úplně* modifying evaluative words, though they are less frequent than the occurrences of evaluative words after *naprosto*.

Úplně seems to combine easily with descriptive adjectives expressing a neutral quality (18).

- (18) *Je úplně svěží a čeká.* ‘She is completely fresh and waiting.’ (SYN2020)

Last, but not least, *úplně* appears in the meaning of “sort of”, “almost” (19).
(19) *Úplně se mi sbíhají sliny.* ‘My mouth is almost watering.’ (ONLINE)

5 CONCLUSIONS

We have carried out the data annotation and analysis of selected intensifiers in order to check out whether human annotators are able to provide consistent and reliable disambiguation decisions based on the available definitions of the adverb and particle categories. Also, because a detailed explanation of the different uses of intensifiers, supported by strong evidence from real data, is usually missing in the current grammars, we wanted to identify important features and properties of the intensifier uses in context that would possibly help shed light on the disambiguation process.

The three intensifiers investigated do not behave in the same way in the context. *Absolutně* is extremely likely to appear in negative contexts, whereas *úplně* acquires a specific meaning with negation in its scope. The annotation suggests that the shift from understanding the intensifier as an adverb to considering it a particle is most advanced with *absolutně*, while *naprosto* is still interpreted in the sense of an adverb, and *úplně* maintains both interpretations. This may also explain the fact that the inter-annotator agreement was the lowest with *úplně*.

Only a few, syntactically restricted types of contexts, allow a definite agreement on the word class assignment, such as intensification of nouns and syntactic nouns, above all totalizing or negative pronouns.

The word class affiliation of intensifiers affecting evaluative expressions is rather unconditioned by the morphological and syntactic interpretation of the affected word. In many cases, the decision is probably driven by semantic factors, above all by the subjective interpretation of its semantics as scalable or non-scalable. This impact of subjective evaluation is a serious factor hindering the efforts to formalize the disambiguation task for automatic analysis.

ACKNOWLEDGEMENTS

This paper has been in part supported by the LINDAT/CLARIAH-CZ project funded by Ministry of Education, Youth and Sports of the Czech Republic (project LM2018101).

This paper has been in part supported by the Ministry of Education of the Czech Republic, through the project Czech National Corpus, no. LM2018137.

References

- [1] Křen, M. et al. (2020). SYN2020: reprezentativní korpus psané češtiny. ÚČNK FF UK, Praha. Accessible at: <http://www.korpus.cz>.

- [2] Hajič, J. et al. (2020). MorfFlex CZ 2.0. Data/software, LINDAT/CLARIAH-CZ digital library, Czech Republic. Accessible at: <http://hdl.handle.net/11234/1-3186>.
- [3] Filipec, J. et al. (2003). Slovník spisovné češtiny pro školu a veřejnost. Praha. [SSČ].
- [4] Čechová, M. (2000). Čeština – Řeč a jazyk. Praha.
- [5] Komárek, M. et al. (1986). Mluvnice češtiny 2. Praha. [MČ2].
- [6] Štícha, F. et al. (2018). Velká akademická gramatika spisovné češtiny. Praha. [VAGSČ].
- [7] Karlík, P. et al. (1995). Příruční mluvnice češtiny. Praha. [PMČ].
- [8] Mathesius, V. (1947). Zesílení a zdůraznění jako jevy jazykové. In Čeština a obecný jazykozpyt, pages 203–223, Praha.
- [9] Vondráček, M. (1999). Příslovce a částice – hranice slovního druhu. Naše řeč 82, pages 72–78.
- [10] Šimková, M. (2002). Příslovky a částice *celkom, úplně – vůbec*. In Varia 9, pages 325–328, Bratislava.
- [11] Havránek, B. et al. (1960–1971). Slovník spisovného jazyka českého. Praha. [SSJČ].
- [12] Akademický slovník současné češtiny. (2017–2020). Praha. [ASSČ]. Accessible at: <https://slovníkcestiny.cz>.
- [13] Cvrček, V., and Procházka, P. (2020). ONLINE_ARCHIVE: monitorovací korpus internetové češtiny. ÚČNK FF UK, Praha. Accessible at: <http://www.korpus.cz>.
- [14] Kopřivová, M. et al. (2017). ORTOFON, verze 1 z 2. 6. 2017. ÚČNK FF UK, Praha. Accessible at: <http://www.korpus.cz>.
- [15] Karlík, P. (2017). Adjektivum. In P. Karlík et al. (eds.), CzechEncy – Nový encyklopedický slovník češtiny. Accessible at: <https://www.czechency.org/slovník/ADJEKTIVUM>.