



**HAL**  
open science

## Definite associative anaphora in informal spoken Czech

Jan Dvořák

► **To cite this version:**

Jan Dvořák. Definite associative anaphora in informal spoken Czech. *Naše řeč*, 2023, 106, pp.87 - 109. 10.58756/n21062301 . hal-04369267

**HAL Id: hal-04369267**

**<https://hal.science/hal-04369267v1>**

Submitted on 4 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Definite associative anaphora in informal spoken Czech: A corpus-based study<sup>1</sup>

---

Jan DVORÁK | Université Toulouse-II-Jean-Jaurès

The present study, couched within the framework of Löbner's Concept Types and Determination theory (CTD) and relying both on corpus data and the questionnaire method, attempts to provide some evidence for the claim that there is a growing tendency in contemporary informal spoken Czech to use the emerging definite article *ten* with definite associative anaphora (DAA). Just like its Western Slavic cognates, the distance-neutral demonstrative *ten* appears to manifest characteristics typical of definite articles across languages (cf. Ortmann, 2014; Czardybon, 2017; Dvořák, 2020). One of these characteristics is the spreading of *ten* to contexts situated between pragmatic and semantic definiteness on Löbner's definiteness scale (Löbner, 1985; 2011). DAA is part of these contexts. However, as the present study shows, marked differences exist between the three sub-types of DAA as defined by Löbner with regards to their willingness to accept *ten*. These are, respectively, the "part-whole," the "relational" and the "situational" sub-type. Other factors must also be taken into account, such as the speaker's emotional involvement and competing interpretations of the occurrence of *ten*.

**Key words:** corpus linguistics, definite associative anaphora, emerging definite article, grammaticalization, informal spoken Czech, *ten*

**Klíčová slova:** gramatikalizace, korpusová lingvistika, neformální mluvená čeština, *ten*, určitá asociální anafora, vznikající určitý člen

## 1 Theoretical background<sup>2</sup>

The Concept Types and Determination theory (CTD), formulated by Löbner (1985; 2011), is a semantic approach to definiteness. In this theoretical framework, definiteness is considered a semantic property of certain, namely individual and functional, nouns. There exist two other types – sortal and relational nouns. These display an a priori indefinite functioning. In addition to the definite/indefinite distinction, another distinction is crucial to CTD – the one between relationality and non-relationality. Thus, relational and functional nouns, contrary to sortal and individual nouns,

---

<sup>1</sup> I would like to thank here Michal Křen from the Institute of the Czech National Corpus for kindly providing me with the transcriptions of the texts of ORTOFON v1, as well as Matthieu Quignard from the CACTUS research group at the ENS de Lyon for helping me to convert the texts into the textometric software TXM. My thanks also go to my ten consultants for Czech.

<sup>2</sup> This section is, to a large extent, taken and adapted from Dvořák (2020, pp. 297–299).

contain an inherent feature of relationality [+R]. Yet there is a substantial difference between relational and functional nouns. Functional nouns are nouns “[...] involving one argument: the possessor. The value for a given argument, in a given context of utterance, constitutes the uniquely determined referent of the noun” (Löbner, 2011, p. 282). Terms like *author* and *perpetrator* are good representatives of this class. Typical examples of relational nouns are the kinship terms *uncle* and *cousin*. For functional nouns, the possessor (anchor) suffices to trigger denotative uniqueness. As for relational nouns, uniqueness does not rely solely on the possessor but also on contextual information. Many nouns can give rise to more than one concept type: the noun *church*, for example, gives rise to a functional concept if we talk about the church of a particular place and if the place takes the form of an explicit or implicit possessive argument (cf. *the church of my neighbourhood*). But it can also yield an individual concept if no argument is present (cf. saying *The church was closed yesterday* when talking about the church of one’s neighbourhood, identifiable by the hearer independently of any associative link).<sup>3</sup> Switching between such meaning variants permanently established within the language system is referred to as level 0 shifts by Löbner (2011).<sup>4</sup> Finally, the noun can also yield a sortal concept, as in *Lyon boasts many splendid churches*.<sup>5</sup> In this case, we are dealing with phenomena called level 1 shifts by Löbner (2011), whereby semantic variants are regularly derived from lexicalized meanings but are not lexicalized since they are too volatile.

Functional and individual nouns constitute semantic definiteness. Inherently indefinite nouns (sortal and relational) can also lend themselves to definite uses when directly drawing on the context. This phenomenon is referred to as pragmatic definiteness. In their definite functioning, sortal and relational nouns have undergone a level 2 shift. This can be seen in the case of *my cousin*, roughly paraphrasable as *the cousin of mine I told you about*. The progression from pragmatic to semantic definiteness can be captured on a scale. This scale contains a buffer zone corresponding to the phenomena at the interface of both types of definiteness. Definite associative anaphora (DAA) is part of these phenomena. The scale below was proposed in Dvořák (2020).

The borderline cases between pragmatic and semantic definiteness (in bold in the scale) include 1) sortal and relational nouns switched into a definite use via the presence of a clause or a prepositional phrase, 2) the different types of DAA and

---

<sup>3</sup> Individual nouns “assign a unique referent to every appropriate context of utterance” (Löbner, 2011, p. 281).

<sup>4</sup> These variants are, in fact, regarded as separate lexicalized meanings of the noun by Löbner.

<sup>5</sup> Sortal nouns make up sortal concepts. “In a given context of utterance, there may be zero, one or more entities the noun denotes.” (ibid., p. 280)

deictic sortal nouns (*Can you see that car?*) < anaphoric sortal nouns (*A man came in. This man looked familiar to me*) < **sortal nouns with complements establishing uniqueness** (*the woman whom I married*) < **situational DAAs** (*Yesterday, I went to the cinema but the film was a flop*) < **relational DAAs** (*This novel is great. I know the author personally*) < **complex individual concepts** (*the biggest rainforest in the world*) < **part-whole DAAs** (*I have bought a new car recently. The engine is already broken*) < lexical individual and functional nouns (*the Prime Minister, my mother*) < proper names (*Jack*) < personal pronouns (*I, me*)

**Figure 1:** Definiteness scale based on Czardybon's 2017 model and reflecting work with Czech data in Dvořák (2020, p. 299).

3) sortal and relational nouns made definite via the presence of modifiers such as ordinals and superlatives. As frequently pointed out in the literature (cf. Himmelmann, 1997; Löbner, 2011; Carlier – De Mulder, 2011; Hofherr – Zribi-Hertz, 2014), the contexts belonging to the buffer zone are particularly relevant to the assessment of the grammaticalization process in which adnominal demonstratives turn into definite articles. Indeed, it is generally assumed that emerging definite articles spread from the domain of pragmatic definiteness (to the left of the buffer zone) to the domain of semantic definiteness (to the right of the buffer zone), and that the grammaticalization process is considered accomplished once the former demonstrative systematically appears in contexts of semantic definiteness.<sup>6</sup> Yet entering the contexts situated between the two types of definiteness and becoming common in these contexts are key moments in the process, and a demonstrative which has reached this stage can be regarded as advanced on its way to becoming a definite article. The reason for this is the fact that genuine demonstratives are banished from these contexts.

Dvořák (2020) provided some additional evidence for the generally shared assumption that informal (spoken) Czech is developing a definite article from the adnominal *ten* (Mathesius, 1926; Vey, 1946; Krámský, 1972; Meyerstein, 1972; Orlandini, 1981; Adamec, 1983; Berger, 1993; Trovesi, 2002; Souček, 2011; Machač – Zíková, 2014; Czardybon, 2017; Zíková, 2017). The present study, using the same corpus data, intends to delve further into the topic by focusing on the presence of *ten* with DAA.

<sup>6</sup> The scale also reflects the tendency of the represented expressions to combine with definite articles. Thus, almost all languages stop using definite articles from a certain point in cases of semantic definiteness, where uniqueness is so obvious that its marking appears as redundant (with proper nouns, typically).

## 2 Definite associative anaphora

In the case of DAA, the definiteness of the head noun is triggered by “a hidden link or anchor which has to be introduced earlier” (Löbner, 1998, p. 1). Given the fact that the anchor needs to be retrieved from the context, Löbner (2011) makes the case that the head noun is subject to a level 2 semantic shift. DAA is situated between pragmatic and semantic definiteness, since it combines properties of both. This hybrid status is summed up by the following quote from Ortmann (2014, p. 309):

DAAs combine properties of pragmatic uniqueness (by virtue of anaphoricity) and semantic uniqueness (by virtue of involving an FN [= a functional noun]). It is therefore natural for there to be considerable variation in the use of articles.

Schwarz (2009), who uses Clark’s term “bridging” (1975), distinguishes two types of DAA, one based on a part-whole link and the other based on a simple relationship between the head noun and the anchor, as in (2). In the part-whole type, the entity denoted by the head noun is part of the entity denoted by the anchor, as in (1):

- (1) We arrived in the village. **The church** was a sumptuous baroque edifice. (invented example)
- (2) When I got on the bus, **the driver** didn’t even say hello to me. (invented example)

Löbner then adds to this distinction a third type, relying on a situational link between the head noun and the anchor. This last – situational – type is not explicitly mentioned in Löbner’s paper on DAA (1998), but Czardybon (2017, p. 66) reports that its existence was communicated to him by the author personally. As an example of situational DAA, see the sentence in (3), taken from Czardybon (2017, p. 71):

- (3) After the cinema they carried on talking about **the film**.

One substantial property differentiates the first two types from the third one. While the head of a part-whole or a relational DAA is an inherently relational functional noun, the head of a situational DAA is a non-relational noun shifted into a relational concept. Thus, the noun *film* in (3) does not contain an inherent sense of relationality [+R], but only acquires it as the result of a level 1 semantic shift:

Both types of DAAs have in common that the head of the DAA is represented by an inherently relational noun. According to Löbner (p.c.), there are DAAs with a [+R] noun as head [...] and ‘situational DAAs’ with a [-R] noun as head. (Czardybon, 2017, p. 66)

The fact that the three sub-types of DAA tend to differ with regard to definiteness marking has been commented on both by Schwarz and by Czardybon. Schwarz points out that in German, where the definite article *der* can be contracted with certain prepositions, the contracted form is preferred (if not regarded as the only correct one) in the realm of part-whole DAAs. This contrasts with the relational type, with which the non-contracted form is preferred (Schwarz, 2009, pp. 34–40). This supports the claim that the part-whole type is closer to pure semantic definiteness than

the relational type. Concerning the situational type, data suggest it is even further away from the pole of semantic definiteness than the relational type. In his comparative overview of definiteness marking in Slavic languages, Czardybon makes the case that in Czech, only the situational DAA allows for the presence of *ten*:

The Czech data shows that situational DAAs can be considered as a separate class of DAAs since only here do we find the possibility of the presence of *ten* in contrast to part-whole and relational DAAs with which *ten* is absent in Czech. (Czardybon, 2017, p. 90)

I believe to have already refuted – at least partly – this claim in Dvořák (2020, p. 315), where I concluded that the analyzed data contained 14 occurrences of *ten* of the relational type. In the present study, even more evidence will be provided for the assertion that the article-like *ten* can appear with the other types of DAA as well, even though the status of *ten* is rather difficult to evaluate when it comes to part-whole DAAs.

### 3 Corpus study

Dvořák (2020) used ORTOFON v1, a balanced corpus of informal spoken Czech, as the major source of data.<sup>7</sup> The assumption underlying both studies is that the spoken variety of Czech is the one where the grammaticalization process of *ten* reaches its highest stages. This assumption is supported by empirical data: in their frequency dictionary of Czech, Čermák and Křen (2004) write that *ten* (including its adnominal as well as pronominal uses) is the 6th most frequent lemma in a corpus consisting of written Czech only; in 2011, the same authors published another frequency dictionary (Čermák – Křen, 2011), where informal spoken language was taken into account (the spoken sub-corpus made up one fourth of the data). As a direct consequence of this, the lemma *ten* got to the position of the 3rd most frequent lemma. In the meantime, *Frekvenční slovník mluvené češtiny* (The Frequency Dictionary of Spoken Czech) was published by a team of authors led by Čermák (2007). This dictionary relies exclusively on informal and semi-formal spoken language data collected in Prague.<sup>8</sup> Its conclusion is striking, as *ten* appears in the 1st position.

---

<sup>7</sup> The corpus was created by the Institute of the Czech National Corpus in Prague and issued in 2017. The data was collected in the period stretching from 2012 to 2017. It contains a total of 1,236,508 positions (1,014,786 orthographic words) from 624 different speakers from all regions of the Czech Republic in 332 recordings. The corpus web page reads: “ORTOFON is also the first corpus to be fully balanced regarding all the basic sociolinguistic speaker categories (gender, age group, level of education and region of childhood residence). The corpus is lemmatized and morphologically tagged in the same manner as the ORAL corpus, the transcription is linked to the corresponding audio track.” (Kopřivová et al., 2017)

<sup>8</sup> In the period between 1988 and 1996. The ranking of the lemmas followed absolute frequency alone. Both adnominal and pronominal occurrences were taken into account.

### 3.1 Method

The corpus data were annotated in the TXM software (Heiden, 2010). The following query, formulated by Dvořák (2020), was used:

```
[pos="PRO\DEM" & lemme!="takov.* "][pos!="VER|CON|PRE|NOM|PCT"]{0,3}[pos="N.* " & word!="vole"]9
```

“For the annotation and analysis, a sample was created by keeping every tenth occurrence of the initial concordance list. Of this sample, the first 1,000 occurrences of the adnominal *ten* were annotated and analyzed. A system of predefined categories was used, distinguishing between the type and the subtype of the use of *ten* [...]” (Dvořák, 2020, p. 303). Attributing a particular value in the annotation process quite often turned out more complex and delicate than expected, as part of the occurrences of *ten* could be accounted for in two different ways. When this happened, I always tried to retain the more plausible interpretation. Regardless of this, 8.1% of the 1,000-occurrence sample were marked as “ambiguous” (Dvořák, 2021b). Unless stated otherwise, all the examples mentioned in this study are taken from ORTOFON v1.

### 3.2 Results

The sample contained 54 occurrences of DAA. This number is higher than the one given in Dvořák (2020), where it only amounted to 47. To justify this rectification, suffice it to say that the way in which the corpus data was analyzed has evolved to some extent since the publication of the previous study. However, as discussed below, regarding some of the occurrences as cases of DAA is potentially controversial. Among the 54 occurrences, 34 seem to belong to the situational sub-type (33 in Dvořák (2020)), while 18 seem to belong to the relational sub-type (14 in Dvořák (2020); see Table 1 below). Only 2 occurrences may, under certain conditions, be associated with the part-whole type, even though this type was not identified at all in Dvořák (2020).

Even though these results must be considered with caution, the situational type turns out to be the most represented. The difference is statistically significant according to the log-likelihood test. The frequencies of the situational and the relational type and those of the situational and the part-whole type were compared. The module “2 words in 1 corpus” of the Corpus Calculator on the Czech National Corpus

---

<sup>9</sup> Other definite demonstratives were also included in the query so that the relative frequency of all demonstratives could be compared to the relative frequency of *ten* alone. Furthermore, measures had to be taken to eliminate false positives such as the indefinite demonstrative form *takový* (‘such’) and the exclamation *ty vole* (‘dude, man’), where *ty* (‘you’) is a personal pronoun.

DAA sub-type	No. of occurrences
Situational	34
Relational	18
Part-whole	2
<b>Total</b>	<b>54</b>

**Table 1:** Numbers of occurrences of *ten* analyzed as cases of the three sub-types of DAA in the 1,000-occurrence sample from ORTOFON v1.

website was used. In both cases, the two frequencies were related to the whole of the 1,000 occurrences. The score 5.1353 ( $p = 0.02344$ ) was obtained for the first difference and the score 34.9797 ( $p < 0.001$ ) was obtained for the second difference.<sup>10</sup> Moreover, the general picture is only partly in compliance with Czardybon's (2017, p. 90) statement, according to which the situational type alone can appear in colloquial Czech.

### 3.3 Discussion

The sentences (4), (5) and (6) are of the situational type:

- (4) S1: jak . vlastně . děl-á-š to aby  
how MOD do-PRS-2SG that.N.ACC.SG that.CONJ+AUX.COND  
ses  
REFL+AUX.PST.2SG  
proces uzdraven-í . jak to vlastně . děl-á-š?  
process.NOM.SG healing-GEN.SG how that.N.ACC.SG MOD do-PRS-2SG
- S2: no zakukl-i-m se . a ček-á-m až . z *té*  
MOD pupate-FUT-1SG REFL and wait-PRS-1SG until from ten.F.GEN.SG  
*kukl-y* .. se vylih-n-e zdrav-ý člověk  
pupa-GEN.SG REFL emerge-FUT.3SG healthy-M.NOM.SG person.NOM.SG
- 'S1: how do... like how do you do that to... the healing process... how do you actually do that?'
- 'S2: well, I pupate and then I wait until a healthy person emerges **from the pupa**'
- (5) no a pak jsme byl-i ještě na jed-nom výlet-u  
MOD and then AUX be.PST-1PL.M still on one-M.LOC.SG trip-LOC.SG  
no to jsme je-l-i džíp-ama naprosto šílen-á  
MOD AUX go-PST-1PL.M jeep-INS.PL completely crazy-F.NOM.SG

<sup>10</sup> See <<https://www.korpus.cz/calc/>>. The difference is statistically significant on the 0.05 level of significance.



cest-a            teda    a    von    ještě *ten*                    *Maročan*  
road-NOM.SG    MOD    and    he.SG    still ten.M.NOM.SG    Moroccan-NOM.SG  
schválně    prostě    najíždě-l                    na *ty*                    *písečn-ý*                    *dun-y*  
on purpose    MOD    drive-PST.3SG.M    on    ten.F.ACC.PL    sand-F.ACC.PL    dune-ACC.PL  
‘and then we went on a trip... well, we went there in jeeps; a totally crazy road; plus **the Moroccan guy** would... he would really like drive with us on **the sand dunes** on purpose’

- (6) S1: nahoře    nad    tím                    byl-y                    tak    obrovitánsk-ý  
upstairs    above    that.N.INS.SG    be.PST-3PL.F    so    gigantic-F.NOM.PL  
pavučin-y  
cobweb-NOM.PL
- S2: jako že    se    tam    rok-y                    ne-vysáva-l-o                    jo ?  
MOD    REFL    there    year-ACC.PL    NEG-vacuum-PST-3SG.N    MOD
- a) ne    on    ten                    týpek                    říká-l                    že  
no    he.NOM    ten.M.NOM.SG    guy.NOM.SG    say-PST.3SG.M    that  
*ti*                    *pavouc-i*                    to                    uděla-l-i  
ten.M.NOM.PL    spider-NOM.PL    that.N.ACC.SG    make-PST-3PL.M  
za hodin-u  
in hour-ACC.SG
- b) ne    on    ten                    týpek                    říká-l                    že  
no    he.NOM    ten.M.NOM.SG    guy.NOM.SG    say-PST.3SG.M    that  
*pavouc-i*                    to                    uděla-l-i                    za hodin-u  
spider-NOM.PL    that.N.ACC.SG    make-PST-3PL.M    in hour-ACC.SG

‘S1: there were such gigantic cobwebs upstairs’

‘S2: like nobody had vacuum-cleaned there for years, right?’

‘S1: no the guy said **the spiders** had made that within a single hour’

(Dvořák, 2020, pp. 313–314)

When focusing on the situational type, one phenomenon struck me as interesting. Unlike within the other two types, where the anchor establishing the uniqueness of the head noun could be pinpointed as a particular NP, within certain cases of the situational type, the argument took the form of the afore-mentioned situation in its totality. Therefore, retrieving the argument by the receiver of the utterance often relies here on a discourse deixis mechanism (Himmelmann, 1996, p. 224), as can be seen in (3):

- (3) After the cinema they carried on talking about *the film*. They liked *the actors*.  
after the cinema (= after going to the cinema) → the film = the film of the situation of going to the cinema

This analysis also works for (4): ‘the pupa of the pupating situation/process.’ In (6), more interpretation schemes seem possible: ‘the spiders of the particular cobwebs / the spiders of the described situation.’ In (5), the NPs *the Moroccan guy* and *the sand dunes* can be interpreted as the Moroccan guy and the sand dunes of the afore-men-

tioned trip. None of the head nouns in (3) to (6) are inherently functional concepts; the [+R] trait thus results from a level 1 semantic shift.

As regards the relational type, Dvořák (2020) comments on the following example:

- (7) já jsem jí říka-l-a že to už  
 I.NOM AUX her.DAT tell-PST-1SG.F that that.N.ACC.SG already  
 m-á úplně to sam-é když pů-jd-e  
 have-PRS.3SG completely ten.N.ACC.SG same-N.ACC.SG if FUT-go-3SG  
 studovat bakalář-e a bude to mít  
 study.INF bachelor's-ACC.SG and AUX that.N.NOM.SG have.INF  
 takov-ou větš-í váh-u než ta  
 such-F.ACC.SG bigger-F.ACC.SG weight-ACC.SG than ten.F.NOM.SG  
 vošk-a přitom [...] **ta** **náročnost**  
 vocational school-NOM.SG yet ten.F.NOM.SG difficulty.NOM.SG  
 je stejn-á si mysl-í-m jakoby  
 be.PRS.3SG same-F.NOM.SG think-PRS-1SG MOD

'so I told her that in the end it'll be the same thing if she does a bachelor's degree [...] it will sort of have more weight than the vocational school while **the demands** are the same I think hour'

The noun *náročnost* ('difficulty') is an inherently relational-functional noun since it calls for an argument triggering its denotative uniqueness (*the difficulty of something*). Let us now consider these two examples:

- (8) ale já m-á-m můj.. telefon v chodb-ě ..  
 but I.NOM have-PRS-1SG my.M.ACC.SG phone-ACC.SG in corridor-LOC.SG  
 teda pevn-ou link-u.. tak na tu kdybyste  
 MOD fixed-F.ACC.SG line-ACC.SG so on this.F.ACC.SG if.CONJ+AUX.COND.2PL  
 zavola-l to vám řek-n-u **to** **čísl-o**  
 call-PTC.SG.M this.N.ACC.SG you.DAT tell-FUT-1SG ten.N.ACC.SG number-ACC.SG  
 'but I have my phone in the corridor... I mean my landline phone... so if you could call me there? I'll tell you **the number**'

- (9) S1: sam-á modřin-a je [...] tam konečně si vyčt-e ..  
 all-F.NOM.SG bruise-NOM.SG be.PRS.3SG there finally REFL read-FUT.3SG  
 že to je [...] z těch lék-ů  
 that that.N.NOM.SG be.PRS.3SG from ten.M.GEN.PL drug-GEN.PL  
 S2: no to já si vždycky čt-u **ty**  
 MOD I.NOM REFL always read-PRS.1SG ten.M.ACC.PL  
**nežádouc-í** **účink-y**  
 side-M.ACC.PL effect-ACC.PL

'S1: she's all covered in bruises [...] there, she'll finally read that it's because of the drugs'  
 'S2: yeah, well, I always read **the side effects**'

The situation of (8) is analogous to that of (7) – *to číslo* is construed as the number of the previously mentioned landline telephone. In (9), the interpretation is a bit more complex as it relies on a series of inferences. Indeed, the lexical compound *side effects* constitutes a functional concept. However, the argument to be retrieved does not correspond exactly to the noun complex *těch léků* ('of the drugs'), mentioned in the text. This interpretation is excluded by the presence of the adjunct *vždycky* ('always'). In the paradigm of generative grammar, the term "quantifier scope/domain" is frequently used for cases like this (cf. Fodor – Sag, 1982; Schwarz, 2009). Here, the domain of the adjunct extends to all the situations under its scope, i.e., those where the speaker is about to start a drug therapy: *every time, before starting to take a drug, I become acquainted with the side effects of this drug*. Therefore, *ty nežádoucí účinky* ('the side effects') is subject to a distributive reading where its reference is fixed separately for each situation within the quantifier scope. I make the case here that the DAA reading of the relational type is the only possible one. Nevertheless, in order for the way in which language users interpret the utterance to become obvious, spelling out the inferences is required:

The drugs that X takes have side effects. Yet all drugs have side effects. Consequently, every time, before starting to take a drug, I read *the side effects of that drug*.<sup>11</sup>

With a couple of occurrences, I hesitated between the relational and situational type:

- (10) on            totiž        uděla-l                    obrovsk-ou                    dír-u                    do  
 he.NOM    since    make-PST.3SG.M    enormous-F.ACC.SG    hole-ACC.SG    into  
 svět-a                    právě            tím                    že    do    těch  
 world-ACC.SG    precisely    that.N.INS.SG    that    into    ten.F.GEN.PL  
rozvojev-ých                    zem-í                    prodáva-l                    levný-ý  
 developing-F.GEN.PL    country-GEN.PL    sell-PST.3SG.M    cheap-F.ACC.PL  
 bot-y                    typ-u                    těch                    dřevák-ů                    a    nějak-ých  
 shoe-ACC.PL    type-GEN.SG    ten.M.GEN.PL    clog-GEN.PL    and    some-M.GEN.PL  
 takov-ých                    jednoduch-ých                    sandál-ů                    kter-ý                    si  
 such-M.GEN.PL    simple-M.GEN.PL    sandal-GEN.PL    which-M.ACC.PL    REFL  
**ty**                    **lid-i**                    moh-l-i                    dovolit  
 ten.M.NOM.PL    people-NOM    can-PST-3PL.M    afford-INF

'he actually put a dent in the universe precisely by selling cheap shoes into the developing countries, cogs, that kind of simple sandals and such like, that **the people** could afford'

- (11) já            m-á-m ..                    nedůvěr-u ..                    v cel-ou                    naš-í  
 I.NOM    have-PRS-1SG    distrust-ACC.SG    in entire-F.ACC.SG    our-ACC.SG

<sup>11</sup> One of the reviewers also observed that the presence of *ten* in (8) and (9) is supported by the fact that the NP appears in a right dislocation. I thank them for this very relevant observation.

<u>politik-u</u>	ale	hlavně	možná	ani ne tak	v <b>ty</b>
politics-ACC.SG	but	mainly	perhaps	not so much	in ten.M.ACC.PL
<b>lid-i</b>	je	to		tím	co
people-ACC	be.PRS.3SG	that.N.NOM.SG		that.N.INS.SG	what.ACC
mně	podáv-a-jí	médi-a			
me.DAT	present-PRS-3PL	média-NOM.PL			

‘I distrust our entire political representation but, most importantly, not so much **the people** as such, it’s [more] about what the media present to me’

Both in (10) and in (11), the DAA noun is *lidi* (‘people’). Given its general meaning and its vague extension, this noun easily adapts to the context where it is used. However, is it possible to regard this noun as inherently relational or, put in other terms, does a relational reading count among its permanent “meaning variants” (Löbner, 2011, p. 310)? Such a reading seems more or less acceptable in (10), where the noun *lidé* (‘people’) lends itself to the interpretation of *obyvatelé* (‘inhabitants’) – *the people of the developing countries*. On the other hand, it does not seem possible in (11), where *lidé* is understood as *the actors of the political system*. Unlike the former variant, the latter is entirely dependent on discourse and is not inscribed within the system. Therefore, (10) was considered a relational DAA, whereas (11) was assigned to the situational type.

The corpus data also contain one occurrence which can be analysed as a case of relational DAA, even though it is, strictly speaking, cataphoric (the endophoric expression precedes the anchor). Indeed, the inherently relational semantics of *rozdíl* (‘difference’) usually calls for a prepositional argument introduced by *mezi* (‘between’):

(12)	ježiš	uč-i-š	studuj-e-š	češtin-u	tak mus-i-š	vidět
	MOD	teach-PRS-2SG	study-PRS-2SG	Czech-ACC.SG	so must-PRS-2SG	see.INF
	<b>ten</b>	<b>rozdíl</b>	.	<u>kysel-ý</u>	a	<u>navinul-ý</u>
	ten.M.ACC.SG	difference.ACC.SG		sour-M.NOM.SG	and	tangy-M.NOM.SG

‘for God’s sake, you teach you study Czech, so you must feel **the difference**: sour and tangy’

The following two examples may be analysed as representing the part-whole subtype of DAA, even though this analysis is somewhat controversial:

(13)	S1:	nech-á-š	to	bejt ..	to	<u>zrcátk-o</u>
		leave-FUT-2SG	that.N.ACC.SG	be.INF	ten.N.ACC.SG	rear-view mirror-ACC.SG
	...					
	S2:	[v] ponděl-í	se	vyměn-í	<b>ten</b>	
		on Monday-ACC.SG	REFL	change-FUT.3SG	ten.M.NOM.SG	
		<b>kryt</b>	a	bud-e	to	vyřízen-ý
		cover-NOM.SG	and	AUX.FUT-3SG	that.N.NOM.SG	solved-N.NOM.SG

‘S1: you’ll leave the rear-view mirror the way it is?’

...

‘S2: on Monday, I’ll get **the cover** changed and the thing will be solved’

- (14) já jsem se byl-a totiž podívat i  
I.NOM AUX.PST.1SG REFL be.PTCP-SG.F actually see.INF even  
v tom byt-ě .. úplně náhodou v tom  
in ten.M.LOC.SG flat-LOC.SG completely by accident in ten.M.LOC.SG  
jejich nov-ým  
theirs-M.LOC.SG new-M.LOC.SG

...

no m-a-j to jakože .. parádn-í ..  
MOD have-PRS-3PL that.N.ACC.SG MOD fabulous-N.ACC.SG  
jakože uzpůsoben-í *těch* *pokojí-ů* a tudle  
MOD layout-ACC.SG ten.M.GEN.PL room-GEN.PL and MOD

‘cause I went to see the flat completely by accident, I mean their new flat

...

well, it’s really like... fabulous like the layout of **the rooms** and stuff’

In (13), *kryt* (‘cover’) is an inherently relational noun. Moreover, a cover is by definition part of the object it covers. Thus, if (13) is a case of DAA, the only acceptable interpretation appears to be one of the part-whole type. In the light of Czardybon’s (2017, p. 90) conclusion, I first tried to exclude this hypothesis. It would be possible to postulate an emotional reading for (13)<sup>12</sup> either if the referent were first introduced without *ten* and then taken up by an expression with *ten*, or if the entire sentence were emotionally marked as in (14). This would correspond to Šimík’s (2015)<sup>13</sup> “pragmatic discourse anaphora,” as the NP would belong to pure semantic definiteness. Yet this reading does not hold here because the NP occurs only once in the conversation. Furthermore, it seems that a recognitional reading cannot be completely ruled out: ‘the cover we talked about the other day / the type of covers you find on rear-view mirrors.’ In this scenario, *kryt* (‘cover’) would be a sortal noun and its reference would be a matter of pure pragmatic definiteness. Consequently, it would not at all rely on the associative link between the mirror and the cover. I presume that one of the reasons why the recognitional reading is possible is that the associative link is not very solid in the speakers’ universe of discourse. Needless to say, the link is more difficult to circumvent when it is more firmly established in speakers’ minds (cf. the one between a car and its engine). In (13), a recognitional reading is all

<sup>12</sup> Emotional (affective) readings are those where the demonstrative does not (directly) participate in the identification of the referent, since it does not modify the extension of the head noun (Mathesius, 1926; Adamec, 1983; Czardybon, 2017; Dvořák, 2020).

<sup>13</sup> See also Adamec (1983).

the more plausible as speaker 1 reminds speaker 2 of a recent incident when the car of the latter was vandalized. Despite this alternative interpretation, a part-whole DAA reading seems perfectly conceivable.

The same conclusion can be made for (14). The noun *místnosti* ('rooms') indeed has a relational functional meaning among its semantic variants (cf. *the rooms of my flat*). At the same time, its referent has not been introduced into the conversation earlier and the interlocutor is not in a position where he could identify it via shared knowledge. However, given the generally emotional and colloquial tone of the utterance – cf. the two occurrences of the colloquial disjunct *jakože*, the subjective and colloquial adjective *parádní* ('fabulous'), the exclamative sentence modality – *ten* could also mark affectivity, hence representing an emotional use.

#### 4 Questionnaire study

Regardless of its countless advantages, working with language corpora also has its limits. These limits lie, first and foremost, in the very nature of corpus data, the originators of which cannot be reached and questioned about their choices. Furthermore, it is not possible to conclude that a language system does not contain a particular phenomenon by relying solely on the absence of this phenomenon within a particular data set. Last but not least, given the size of the corpus material, it was not conceivable to include in this study all cases of DAA, i.e., also those without *ten*. For all these reasons, I found it judicious to complete the corpus data with a questionnaire distributed to ten native speakers of Czech.

##### 4.1 Participants, material and procedure

The group of consultants consisted of five women and five men. Most of them were born at the beginning of the 1990s, and all except for one are university graduates (of translation studies, history, and sociology). Table 2 is a summary of the consultants' information:

Consultant	Sex	Date of birth
1	F	1992
2	F	1992
3	F	1991
4	F	1991
5	F	1990

Consultant	Sex	Date of birth
6	M	1991
7	M	1991
8	M	1990
9	M	1972
10	M	1992

**Table 2:** Native Czech consultants.

Although the questionnaire had been designed in such a way as to cover a whole range of contexts where *ten* can appear as an emerging definite article,<sup>14</sup> I will limit myself here to the examples related to DAA. The questionnaire comprised an introduction outlining the research objectives and providing the necessary instructions. It particularly emphasized the informal and colloquial character of the context where the sentences appeared, and it explicitly invited the consultants to pick the variant they were likely to use in an everyday informal conversation with friends. The informal character was conveyed by the presence of morphological, syntactic and lexical features typical of colloquial language. For each of the five sentences, the consultants were asked to choose either the variant with *ten* or without *ten*. They were encouraged to leave me any comments they deemed useful:

- (15) Po kin-ě se ještě bavi-l-i o *tom* *film-u*.  
 after cinema-LOC.SG REFL still talk-PST-3PL.M about ten.M.LOC.SG film-LOC.SG  
 Po kin-ě se ještě bavi-l-i o *film-u*.  
 after cinema-LOC.SG REFL still talk-PST-3PL.M about film-LOC.SG  
 ‘After the cinema they talked about **the film**.’ (Czardybon, 2017, p. 90)

- (16) Když jsem veše-l do toho křám-u,  
 when AUX.PST.1SG enter-PTCP-SG.M into ten.M.GEN.SG shop-GEN.SG  
 a) tak se mě *prodavačk-a* hned  
 so REFL me.ACC shop assistant-F.NOM.SG immediately  
 pta-l-a, co chc-i.  
 ask-PST-3SG.F what.ACC want-PRS-1SG  
 b) tak se mě *ta* *prodavačk-a* hned  
 so REFL me.ACC ten.F.NOM.SG shop assistant-F.NOM.SG immediately  
 pta-l-a, co chc-i.  
 ask-PST-3SG.F what.ACC want-PRS-1SG  
 ‘When I entered the shop, **the shop assistant** immediately asked me what I wanted.’  
 (invented example)

- (17) S1: nahoře nad tím byl-y tak obrovitánsk-ý  
 upstairs above that.N.INS.SG be.PST-3PL.F so gigantic-F.NOM.PL  
pavučin-y  
 cobweb-NOM.PL  
 S2: jako že se tam rok-y ne-vysáva-l-o jo ?  
 MOD REFL there year-ACC.PL NEG-vacuum-PST-3SG.N MOD  
 a) ne on ten týpek říká-l že  
 no he.NOM ten.M.NOM.SG guy.NOM.SG say-PST.3SG.M that

<sup>14</sup> The questionnaire was designed as part of the research project of my PhD dissertation.

*ti*                      *pavou-ci*                      to                      uděla-l-i  
 ten.M.NOM.PL      spider-NOM.PL      that.N.ACC.SG      make-PST-3PL.M  
 za hodin-u  
 in hour-ACC.SG

- b) ne on                      ten                      týpek                      říká-l                      že  
 no he.NOM                      ten.M.NOM.SG                      guy.NOM.SG                      say-PST.3SG.M                      that  
*pavou-ci*                      to                      uděla-l-i                      za hodin-u  
 spider-NOM.PL                      that.N.ACC.SG                      make-PST-3PL.M                      in hour-ACC.SG

‘S1: there were such gigantic cobwebs upstairs’

‘S2: like nobody had vacuum-cleaned there for years, right?’

‘S1: no the guy said **the spiders** had made that within a single hour’

(Dvořák, 2020, pp. 313–314)

- (18) Když jsem                      posledně                      nastoupi-l                      v Praze-  
 when AUX.PST.1SG                      last time                      get on-PTCP.SG.M                      in Prague-LOC.SG  
 do autobus-u,  
 into bus-GEN.SG

- a) tak mě *ten*                      řidič                      ani ne-pozdravi-l.  
 so me.ACC ten.M.NOM.SG driver.NOM.SG even NEG-salute-PST.3SG  
 b) tak mě *řidič*                      ani ne-pozdravi-l.  
 so me.ACC driver.NOM.SG even NEG-salute-PST.3SG

‘Last time I got on a bus in Prague, **the driver** didn’t say hello to me.’ (invented example)

- (19) Petr                      si                      koupi-l                      nov-ý                      aut-o.  
 Peter.NOM.SG                      REFL                      buy-PST.3SG.M                      new-N.ACC.SG                      car-ACC.SG  
 a) Hned                      druh-ej                      den                      se                      mu  
 immediately                      second-M.ACC.SG                      day.ACC.SG                      REFL                      him.DAT  
 rozbi-l                      *ten*                      *motor*.  
 break down-PST.3SG.M                      ten.M.NOM.SG                      engine.NOM.SG

- b) Hned                      druh-ej                      den                      se                      mu  
 immediately                      second-M.ACC.SG                      day.ACC.SG                      REFL                      him.DAT  
 rozbi-l                      *motor*.  
 break down-PST.3SG.M                      engine.NOM.SG

‘Peter bought a new car. On the very next day, **the engine** broke down.’

(invented example).

## 4.2 Results and discussion

Sentences (15), (16) and (17) belong to the situational type. In (15) and (17), 8 and 9 consultants respectively chose the option with *ten* (see Table 3 below). On the other hand, only one consultant used *ten* with *prodavačka* (‘shop assistant’) in (16). A possible explanation might reside in the fact that the term *prodavačka* may lend itself to an indefinite reading – there can indeed be several shop assistants in a shop.



One consultant even told me that to her, the presence of *ten* necessarily entails a recognitional reading (Himmelmann, 1996), that is, a reading independent of the discourse context – *the shop assistant, you know which one, the one I told you about*. In (18), where *řidič* (‘driver’) is the driver of the particular bus on which the speaker got to Prague, only two consultants opted for *ten řidič*, which is in compliance with the hypothesis that speakers are more reluctant to use *ten* with the relational type.<sup>15</sup>

Example	Number of consultants having chosen <i>ten</i>
(15) <i>the cinema – the film</i>	8
(16) <i>the shop – the shop assistant</i>	1
(17) <i>such gigantic cobwebs – the spiders</i>	9
(18) <i>a bus – the driver</i>	2
(19) <i>a new car – the engine</i>	0

**Table 3:** Numbers of consultants who chose the option with *ten* in examples (15)–(19).

None of the consultants chose the variant with *ten* in (19), which represents the part-whole type. This was in line with my expectations, as the presence of *ten* seems very unnatural here. Yet what could possibly account for the fact that *ten* is acceptable in (13) and (14) above but not in (19)? Unfortunately, I am, as of now, unable to provide a satisfactory answer to this question, except for the afore-mentioned hypothesis that neither (13) nor (14) are actually cases of (pure) DAA. I would like to thank Olga Nádvořníková, who shared with me the following consideration: should the sentence in (19) be transformed in such a way as to be overtly emotional, *ten* may become acceptable:

- (20) *Si představ, že Petr si koupi-l*  
 REFL imagine.IMP.2SG that Peter.NOM.SG REFL buy-PST.3SG.M  
*nov-ěj bourák, ale hned druh-ěj den*  
 new-M.ACC.SG whip-ACC.SG but immediately second-M.ACC.SG day.ACC.SG  
*se mu rozbi-l ten motor!*  
 REFL him.DAT break down-PST.3SG.M ten.M.NOM.SG engine.NOM.SG  
 ‘*imagine!* Peter bought himself a *new whip*, but on the very next day, *the engine* broke down!’  
 (invented example)

<sup>15</sup> Even though both a shop assistant and a driver practice their profession within a particular stereotypical framework, the noun *driver* is inherently relational since *the driver of the bus* is perfectly natural, contrary to *the shop assistant of the shop*, at least according to my judgement. This notwithstanding, I cannot exclude a possible inherently relational meaning for *shop assistant* either. I also have to admit that this reading could better explain why, in both sentences, *ten* is not chosen by most of the consultants.

In (20), *ten* combines DAA marking with emotional marking, analogously to what happens in certain contexts with “complex individual concepts,” that is, sortal or relational nouns made definite through the presence of a particular class of modifiers (Czardybon, 2017):

- (21) všichn-i      m-a-j              takov-ý              obř-í              plnovous-y  
 all-NOM.PL    have-PRS-3PL    such-M.ACC.PL    giant-M.ACC.PL    beard-ACC.PL
- že jo      chod-i              v maskáč-ich              je              to  
 MOD    walk-PRS-3PL    in camo suit-LOC.PL    be.PRS.3SG    that.N.NOM.SG
- fakt      zuby              ne-m-a-jí              fakt      *tí*  
 really    tooth-ACC.PL    NEG-have-PRS-3PL    really    ten.M.NOM.PL
- nej-věš-í*              *buran-i* ..  
 SUP-bigger-M.NOM.PL    redneck-M.NOM.PL
- ‘they all have sort of giant beards, you see, they wear camo suits, it’s really...  
 they have no teeth, really the worst rednecks’

Dvořák (2020; 2021a) made the case that in view of its abundant presence in colloquial Czech, the emotional value may act as an encouraging factor in the transformation of *ten* into a full-fledged definite article.<sup>16</sup>

Unfortunately, the different degree of acceptability of *ten* with respect to factors such as emotionality, the fact that the questionnaire only focused on two options (the presence or the absence of *ten*) as well as the rather homogeneous composition of the group of consultants also make the questionnaire results rather inconclusive. Indeed, all of these factors seem too important to allow for any general statements.

## 5 Definite associative deixis?

In addition to the three sub-types of DAA, the existence of another phenomenon can be mentioned here, which I venture to call “definite associative deixis,” by analogy with the previously discussed examples. This phenomenon, exemplified by (22), has been theorized about extensively in the literature. It is referred to as “larger situation use based on general knowledge” by Hawkins (1978).

- (22) A person enters a supermarket and asks a member of the staff the following question:  
 Where’s *the frozen food aisle*, please?

Contrary to DAA, where the anchor of the functional concept is provided in the previous discourse, in (22), it is provided by the situational context. Therefore, regard-

<sup>16</sup> For the importance of subjectivity and pragmatic factors at the beginning of the process whereby demonstratives are grammaticalized into definite articles, see for example Epstein (1995), Traugott and Hopper (2003/1993), and Carlier and De Mulder (2011).

less of the impossibility for the speaker to perceive and to identify a particular referent (they cannot see the aisle, hence their question), the NP is definite because of the conventional link between supermarkets and their aisles (there exists a frozen food aisle in every supermarket) – in other words, between a permanently established cognitive frame and its elements. It may strike us as odd that in today's colloquial Czech, the noun is always bare when sentences like (22) are uttered in an emotionally neutral context.<sup>17</sup> This can undoubtedly be explained by the fact that, in comparison to DAA, the retrieval of the anchor relies more on world knowledge (and, consequently, on inferences). Thus, the anchor is even more implicit than in DAA, where it does appear in the discourse. Clearly, this absence of discourse-anchoring brings cases like (22) even closer to the pole of pure semantic definiteness, even though – as has been seen above – implicitness and inferences can also be (largely) at stake in DAA. Therefore, we can make the prediction that *ten* will only become possible in examples like (22) once it has become more or less systematic in all contexts of DAA.

## 6 Conclusion and perspectives

In none of the examples of DAA sentences commented on in this study is the presence of *ten* strictly speaking obligatory, even though it is clearly preferred in certain contexts of the situational sub-type.<sup>18</sup> It may also look quite surprising that within the very same domain of situational DAA, there is a marked proclivity to use *ten* only with some examples, whereas it is almost unanimously avoided in other examples (cf. the very different feedback from my consultants for (15) and (17) on the one hand and for (16) on the other). There is no doubt that *ten* is possible with at least certain cases of the relational sub-type, as attested by the corpus examples (7), (8) and (9). However, both the data and the consultants' feedback suggest that the reluctance to use *ten* is stronger here than in the situational sub-type – 34 and 18 occurrences of the data sample seem to belong to situational DAA and relational DAA respectively<sup>19</sup> (the difference is statistically significant according to the log-likelihood test), and

---

<sup>17</sup> I presented the consultants with a Czech translation of (22). None of them chose *ten*. Similarly to (20), the presence of *ten* cannot but place the sentence in an emotionally marked context (e.g., one where the speaker is irritated by having to repeat their question after obtaining no answer from the staff member).

<sup>18</sup> One of the reviewers objected that in certain cases, the absence of *ten* would make the sentence sound unnatural or would even change its interpretation. We agree with this objection.

<sup>19</sup> Unfortunately, the size of the corpus sample prevented me from taking into account DAA occurrences without *ten* in order to establish the proportion of bare NPs and NPs with *ten* within each sub-category. The questionnaire was an attempt to partly remedy this shortcoming. In my future research, I intend to analyze both phenomena in a corpus of smaller size.

only two consultants chose *ten* with *řidič* ('driver') in (18). As far as the part-whole sub-type is concerned, *ten* seems to be ruled out in many (most?) cases – see the unanimous rejection of *ten* by my consultants in (19). This outcome is in line with predictions based on the CTD approach, which situates the part-whole sub-type closest to the pole of semantic definiteness. The only two occurrences labelled as possible part-whole DAAs in the sample turned out rather controversial in terms of their analysis. We could indeed see that in neither case is the DAA reading the only possible one – while the referent of (13) might be identified independently of its associative link with the rear-view mirror, *ten* in (14) allows for a reading where DAA and emotionality are conflated. The latter statement has been confirmed via (20), overtly displaying the speaker's emotional involvement. At last, I tried to show how a parallel can be drawn between DAA and Hawkins's "larger situation uses based on general knowledge," since in both cases, we have a functional concept in need of being saturated by a contextual argument for its definiteness. However, what I therefore decided to call "definite associative deixis" is even closer to pure semantic definiteness than part-whole DAA, since the argument is never mentioned and must be retrieved from the situation itself.

As regards inferences, I stressed their importance for the entire category of DAA. In general, I am inclined to conclude that inferences tend to play a more crucial role within the situational sub-type, since, as can be notably observed in (4), (6) and (15), the hearer often has to infer here that the argument takes the form of the situation described in its totality. On the other hand, we saw in (9) that some kind of inferring is necessary in certain non-situational DAAs as well. In fact, we can easily conclude that the notion of inference is vital for all the DAAs whose argument is not made explicit in the text:

(23) *Tomorrow, I will ask him what he thinks about that. As soon as I get **the answer**, I'll let you know.*

In (23), *the answer* is interpreted as 'the answer to the question I will pose to him when asking him what he thinks about that.' Thus, the argument (*the question*) is implicit and has to be retrieved from the meaning of the sequence *I will ask him what he thinks about that*. While the question-answer link belongs to the relational sub-type, the one between the process of asking and the question is of the situational sub-type.

In spite of the general statement that the situational type shows the highest affinity for definiteness marking and in spite of what I proposed in order to account for some of the discrepancies in the data – e.g., the possibility of an indefinite reading for *prodavačka* ('shop assistant') in (16), which might partly explain why so few consultants chose *ten* –, I have to admit my current incapacity to fully elucidate the factors underlying the use of *ten* with DAA. As has been stressed in this study, at least

two factors seem to play an important role. On the one hand, it is emotionality, as manifested in (14) and (20). On the other hand, it is the tightness of the associative link between the head noun and the anchor in speakers' universe of discourse, as manifested by the comparison between (13) and (19).<sup>20</sup> This is why, in my future research, I would like to investigate these two factors more thoroughly.

I also have to stress here the limits stemming from the restricted number of the occurrences analyzed. The relatively low presence of DAA within the corpus sample may be related, among other things, to an over-representation of other types of occurrences, in particular those representing emotional uses and recognitional uses (see Dvořák, 2020; 2021b). Having said this, the percentage of DAA featured in the sample (5.4% of all the occurrences) still seems to constitute quite satisfactory evidence of definiteness-marking in the realm of DAA in informal spoken Czech in light of the fact that Czech continues to be regarded as an articleless language (cf. Dryer et al., 2013). In comparison, consider Poesio and Vieira's (1997) analysis of English definite descriptions in a written corpus of newspaper articles. The authors' sample consisted of 1,412 occurrences, and the percentage of those analyzed as DAA varied from 14.42% to 18.55%.<sup>21</sup> No matter how striking the difference, it actually seems quite logical with regards to the fact that English has a full-fledged definite article.

## REFERENCES

- ADAMEC, Přemysl (1983): České zájmeno *ten* a jeho ruské ekvivalenty. In: Vladimír Hrabě – Aleksandra Grigor'jevna Širokova (eds.), *Konfrontační studium ruské a české gramatiky a slovní zásoby* [2. svazek]. Praha: Universita Karlova, pp. 153–170.
- BERGER, Tilman (1993): *Das System der tschechischen Demonstrativpronomina: Textgrammatische und stilspezifische Gebrauchsbedingungen* [unpublished habilitation thesis]. München: Ludwigs-Maximilians-Universität.
- CARLIER, Anne – DE MULDER, Walter (2011): The grammaticalization of definite articles. In: Bernd Heine – Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*. Oxford – New York, NY: Oxford University Press, pp. 522–534.  
<<https://doi.org/10.1093/oxfordhb/9780199586783.013.0042>>.
- CLARK, Herbert H. (1975): Bridging. In: Roger Schank – Bonnie L. Nash-Webber (eds.), *Theoretical Issues in Natural Language Processing*. New York, NY: Association for Computing Machinery, pp. 169–174.

---

<sup>20</sup> When comparing (13) to (19), it seems quite clear to me that the “obviousness” of the link is not grounded in language itself but in speakers' extralinguistic experience. From a strictly semantic point of view, it seems as natural for rear-view mirrors to have covers as it is for cars to have engines: what makes a substantial difference here is the degree of speakers' familiarity with both facts.

<sup>21</sup> The result variation was due to the presence of three different annotators.

- CZARDYBON, Adrian (2017): *Definiteness in a Language without Articles – A Study on Polish*. Düsseldorf: Düsseldorf University Press. <<https://doi.org/10.1515/9783110720426>>.
- ČERMÁK, František et al. (2007): *Frekvenční slovník mluvené češtiny*. Praha: Karolinum.
- ČERMÁK, František – KŘEN, Michal (2011): *A Frequency Dictionary of Czech: Core Vocabulary for Learners*. London – New York, NY: Routledge.
- ČERMÁK, František – KŘEN, Michal et al. (2004): *Frekvenční slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- DRYER, Matthew S. (2013): Czech. In: Matthew S. Dryer et al., *The World Atlas of Language Structures Online* [online]. Leipzig: Max Planck Institute for Evolutionary Anthropology. Last Accessed: March 15, 2023. <[https://wals.info/languoid/lect/wals\\_code\\_cze](https://wals.info/languoid/lect/wals_code_cze)>.
- DVOŘÁK, Jan (2020): The emerging definite article *ten* in (informal spoken) Czech: a further analysis in terms of semantic and pragmatic definiteness. *Naše řeč*, 103(4), pp. 297–319.
- DVOŘÁK, Jan (2021a): L'emploi du démonstratif *TEN* avec l'adjectif superlatif en tchèque: entre simple marquage de la définitude et effets pragmatiques. *Scolia*, 35, pp. 13–31. <<https://doi.org/10.4000/scolia.1823>>.
- DVOŘÁK, Jan (2021b): *Les emplois adnominaux des démonstratifs tchèque et français dans la langue orale informelle: étude de sémantique référentielle contrastive* [unpublished doctoral thesis]. Lyon: École normale supérieure de Lyon.
- EPSTEIN, Richard (1995): L'article défini en ancien français: l'expression de la subjectivité. *Langue française*, 107(1), pp. 58–71.
- FODOR, Janet Dean – SAG, Ivan (1982): Referential and quantificational indefinites. *Linguistics and Philosophy*, 5(3), pp. 355–398.
- HAWKINS, John A. (1978): *Definiteness and Indefiniteness*. London: Croom Helm.
- HEIDEN, Serge (2010): The TXM platform: building open-source textual analysis software compatible with the TEI encoding scheme. In: Ryo Otoguro – Kiyoshi Ishikawa – Hiroshi Umemoto – Kei Yoshimoto – Yasunari Harada (eds.), *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation: Institute for Digital Enhancement of Cognitive Development*. Sendai: Institute of Digital Enhancement of Cognitive Processing, Waseda University, pp. 389–398.
- HIMMELMANN, Nikolaus P. (1996): Demonstratives in narrative discourse: a taxonomy of universal uses. In: Barbara A. Fox (ed.), *Studies in Anaphora*. Amsterdam – Philadelphia, PA: John Benjamins. <<https://doi.org/10.1075/tsl.33.08him>>.
- HIMMELMANN, Nikolaus P. (1997): *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Berlin: De Gruyter. <<https://doi.org/10.1515/9783110929621>>.
- HOFHERR, Patricia Cabredo – ZRIBI-HERTZ, Anne (eds.) (2014): *Crosslinguistic Studies on Noun Phrase Structure and Reference*. Leiden: Brill. <<https://doi.org/10.1163/9789004261440>>.
- KRÁMSKÝ, Jiří (1972): *The Article and the Concept of Definiteness in Language*. The Hague: Mouton. <<https://doi.org/10.1515/9783110886900>>.
- LÖBNER, Sebastian (1985): Definites. *Journal of Semantics*, 4(4), pp. 279–326. <<https://doi.org/10.1093/jos/4.4.279>>.
- LÖBNER, Sebastian (1998): Definite associative anaphora. In: Simon Philip Botley (ed.), *Approaches to Discourse Anaphora: Proceedings of DAARC96 – Discourse Anaphora and Resolution Colloquium, Lancaster University, July 17th–18th 1996*. Lancaster: University Centre for Computer Corpus Research on Language.

- LÖBNER, Sebastian (2011): Concept types and determination. *Journal of Semantics*, 28(3), pp. 279–333. <<https://doi.org/10.1093/jos/ffq022>>.
- MACHAČ, Pavel – ŽÍKOVÁ, Magdalena (2014): Variability in phonetic realization of the demonstrative *ten* in terms of its informational relevance in the sentence. In: Ludmila Veselovská – Markéta Janebová (eds.), *Nominal Structures: All in Complex DPs*. Olomouc: Palacký University, pp. 98–109.
- MATHESIUŠ, Vilém (1926): Přívlastkové *ten, ta, to* v hovorové češtině. *Naše řeč*, 10(2), pp. 39–41.
- MEYERSTEIN, Zlata P. (1972): Czech deictics: pronouns and articles? *Linguistics*, 10(91), pp. 17–30. <<https://doi.org/10.1515/ling.1972.10.91.17>>.
- ORLANDINI, Anna (1981): Wesen und Entwicklung des Artikels vom Lateinischen zu den Romanischen Sprachen. *Indogermanische Forschungen*, 86(1), pp. 223–247. <<https://doi.org/10.1515/9783110243284.223>>.
- ORTMANN, Albert (2014): Definite article asymmetries and concept types: semantic and pragmatic uniqueness. In: Thomas Gamerschlag – Doris Gerland – Rainer Osswald – Wiebke Petersen (eds.), *Frames and Concept Types: Applications in Language and Philosophy*. Basel: Springer, pp. 293–321. <[https://doi.org/10.1007/978-3-319-01541-5\\_13](https://doi.org/10.1007/978-3-319-01541-5_13)>.
- ORTOFON v1 [online; version 1 from December 28, 2017] (2017). Praha: Ústav Českého národního korpusu FF UK. Last accessed: March 26, 2023. <[www.korpus.cz](http://www.korpus.cz)>.
- POESIO, Massimo – VIEIRA, Renata (1997): A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2), pp. 1–46. <<https://doi.org/10.48550/arXiv.cmp-lg/9710007>>.
- SCHWARZ, Florian (2009): *Two Types of Definites in Natural Language* [unpublished doctoral thesis]. Amherst, MA: University of Massachusetts. <<https://doi.org/10.7275/1077454>>.
- ŠMÍK, Radek (2015): On pragmatic demonstratives: the case of pragmatic discourse anaphora in Czech. *Proceedings of Sinn und Bedeutung*, 20, pp. 640–657.
- SOUČEK, Milan (2011): Má čeština určitý člen? Komparativní analýza funkcí určitého členu *the* v angličtině a zájmen *ten, ta, to* v češtině. In: František Čermák (ed.), *Korpusová lingvistika Praha 2011: 1, InterCorp*. Praha: Nakladatelství Lidové noviny, pp. 64–72.
- TRAUGOTT, Elizabeth C. – HOPPER, Paul J. (2003/1993): *Grammaticalization*. Cambridge: Cambridge University Press.
- TROVESI, Andrea (2002): Určenost nominální skupiny v češtině, horní lužické srbštině a slovinštině. In: Alena Krausová – Markéta Slezáková – Zdeňka Svobodová (eds.), *Setkání s češtinou (Sborník z konference Setkání s češtinou konané v Praze 6.–7. září 2001)*. Praha: Ústav pro jazyk český Akademie věd ČR, pp. 129–136.
- VEY, Marc (1946): *Morphologie du tchèue parlé*. Paris: C. Klincksieck.
- ŽÍKOVÁ, Magdalena (2017): *Gramatikalizační potenciál anaforické funkce lexému ten v mluvených narativěch* [unpublished doctoral thesis]. Praha: Filozofická fakulta UK.

### Určitá asociační anafora v neformální mluvené češtině: korpusová studie

Studie vychází z rámce „Concept Types and Determination theory“ (CTD) S. Löbnera a opírá se jak o korpusová data, tak o metodu dotazníkového šetření. Snaží se ověřit tvrzení o rostoucím užívání vznikajícího určitého členu *ten* v rámci určité asociační anafory (DAA) v současné neformální mluvené češtině. Demonstrativum *ten*, jež je neutrální z hlediska vyjadřování vzdálenosti, vykazuje stejně jako jeho kognáty v jiných západoslovanských jazycích vlastnosti, které jsou pro určité členy

typické napříč jazyky (srov. Ortmann, 2014; Czardybon, 2017; Dvořák, 2020). Jedna z těchto vlastností je pronikání výrazu *ten* do kontextů nacházejících se na pomezí pragmatické a sémantické určitosti na Löbnerově škále určitosti (Löbner, 1985; 2011). DAA je součástí těchto kontextů. Studie však zároveň ukazuje, že s ohledem na ochotu akceptovat demonstrativum *ten* existují mezi třemi podtypy DAA značné rozdíly. Jedná se o podtyp „část–celek“ a relační a situační podtyp/podtypy. Studie rovněž dokládá, že je nutné zohlednit i další faktory, např. citové zaujetí mluvčího, a více možností, jak interpretovat daný výskyt výrazu *ten*.

Université Toulouse 2 – Jean Jaurès, Département Lettres modernes, Cinéma et Occitan  
Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE)  
5, allées Antonio Machado, 31058 Toulouse cedex 9, France  
jan.dvorak@univ-tlse2.fr



Tento článek je volně přístupný na základě licence *Creative Commons Uvedte původ 4.0 Mezinárodní veřejná licence*, která umožňuje používání, šíření a reprodukci článku za podmínky, že původní článek bude řádně citován. Licenční podmínky jsou dostupné zde: <https://creativecommons.org/licenses/by/4.0/legalcode.cs>.

This is an open access article under the terms of the *Creative Commons Attribution 4.0 International Public License*, which permits use, distribution and reproduction in any medium, provided the original article is properly cited. The license terms are available here: <https://creativecommons.org/licenses/by/4.0/legalcode>.