



HAL
open science

The emerging definite article ten in (informal spoken) Czech: a further analysis in terms of semantic and pragmatic definiteness

Jan Dvořák

► To cite this version:

Jan Dvořák. The emerging definite article ten in (informal spoken) Czech: a further analysis in terms of semantic and pragmatic definiteness. *Naše řeč*, 2019. hal-03704697

HAL Id: hal-03704697

<https://hal.science/hal-03704697v1>

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

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/361309249>

The emerging definite article ten in (informal spoken) Czech: a further analysis in terms of semantic and pragmatic definiteness

Article · November 2019

CITATION

1

1 author:



Jan Dvorak

Sorbonne Université

6 PUBLICATIONS 2 CITATIONS

SEE PROFILE

The emerging definite article *ten* in (informal spoken) Czech: a further¹ analysis in terms of semantic and pragmatic definiteness²

Jan DVOŘÁK | Institut d’Histoire des Représentations et des Idées dans les Modernités, École normale supérieure de Lyon

The present study, falling within the framework of the Concept Types and Determination theory (CTD) and relying upon corpus data, attempts to provide further evidence for the claim that Czech, especially its informal spoken variety, is developing a definite article from the distance-neutral demonstrative *ten* in adnominal uses. The CTD theory has proved its utility for studying emerging definite articles in Western Slavic languages in the works of Adrian Czardybon and Albert Ortmann. At its core lies the distinction between the so-called “pragmatic” and “semantic” definiteness. It is generally assumed that emerging definite articles spread from the former to the latter, and the grammaticalization process is considered accomplished once the former demonstrative systematically appears in contexts of semantic definiteness. This study applies the distinction, made by Löbner, to a corpus sample of 1,000 occurrences of the adnominal *ten*, many of which appear to manifest characteristics typical of definite articles across languages.

Key words: corpora, definite article, demonstrative, grammaticalization, informal spoken Czech

1 Introduction

The Concept Types and Determination theory (CTD) was developed by Löbner (1985; 2011). It is articulated on a semantic basis wherein definiteness is understood as uniqueness of reference and is said to be a default semantic property of certain nouns, i.e. the so called “individual” and “functional” nouns, while two further concept types of nouns – i.e. “sortal” and “relational” – are a priori predestined to non-

¹ The present study follows on chiefly from Czardybon (2017).

² Here, we would like to thank several people without whom the present study could not have been carried out. First of all, we would like to express our gratitude to Carla Bombi and Radek Šimik, the organizers of the workshop *Sorting out the Concepts behind Definiteness*, held as part of 41st Annual Conference of the German Linguistic Society in March 2019, whose theme inspired the study. Then we would like to thank Michal Křen from the Institute of the Czech National Corpus for kindly providing us with the transcriptions of the corpus ORTOFON v1 as well as Matthieu Quignard from the CACTUS research group for helping us with converting these texts into the textometric software TXM and with the subsequent processing of the annotated data. Our thanks also go to Sebastian Löbner and Adrian Czardybon for their responsiveness to our emails and their willingness to discuss their work with us.

unique reference. Individual nouns are, in Löbner's (2011, p. 281) words, nouns whose "meanings are individual concepts that assign a unique referent to every appropriate context of utterance". An example of an individual noun is *town hall* since it has a unique referent for the inhabitants of a particular town. Functional nouns are nouns whose "meanings are functional concepts, 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). Examples of functional nouns are the kinship terms *mother* and *father*.³ Sortal nouns are "unary predicate terms [...]. Their meanings are sortal concepts. In a given context of utterance, there may be zero, one or more entities the noun denotes" (ibid., p. 280). The noun *individual* would be an example. "Relational nouns are binary predicate terms [...]. Their meanings are binary relational concepts, involving a further argument in addition to the referential argument" (ibid., p. 281). Typical examples of relational nouns are the kinship terms *brother* and *sister*. Unlike functional nouns, where the possessor itself triggers a unique reference, in the case of relational nouns, in order for unique reference to be established, the presence of a certain form of explicit contextual information is compulsory. Functional and individual nouns make up "semantic definiteness". The term "pragmatic definiteness" then covers those cases where an inherently non-unique noun – that is a sortal or a relational one – is uniquely referential due to a situational or discursive context. The two situations correspond to deixis and anaphora respectively.

A crucial aspect of Löbner's theory is that many nouns are basically ambiguous insofar as they can belong to several concepts at the same time. For instance, the noun *house* can be used as a functional concept in *my house* (the presence of the one possessor argument suffices, within the given context, to make the noun refer uniquely) and at the same time as a sortal concept in *There are many nice houses in Bremen*. Here, both concepts coexist latently within the semantics of the term *house* and make up what Löbner calls the noun's semantic interpretations (or variants). The choice of one or the other depends on the particular use of the term and is accounted for in terms of "level zero semantic shifts". Two other types of shifts exist: "level 1", wherein semantic variants are regularly derived from lexicalized meanings but are not themselves lexicalized. Typically, this phenomenon includes nouns

³ An important point to be made is that even in the case of individual and functional nouns, the referent "depends on the context of utterance, including a time index, a location index and a constellation of facts [...]. The range of indices for which an individual noun refers to a particular entity may vary widely. Some have almost globally uniform reference (*US president, Ghana*), while others like *mosque, station, exit* and so on have a particular referent only in smaller domains" (Löbner, 2011, p. 284). The latter case constitutes what Gerland and Horn (2010) call "permanently established individual concepts", i.e. nouns which only refer to individual concepts within a shared restricted setting, e.g. *the dog, the kitchen, and the garbage bin* within a particular family setting.

such as *car*, *flat* or *toothbrush*, which tend to have only one referent within a particular situation of use. Yet these functional interpretations are volatile and are so far not coded as permanent semantic variants in the lexicon. And “level 2” shifts, which “draw on particular contextual information” (Löbner, 2011, p. 310). An example is the sortal noun *man*, which is given an individual interpretation via a previous mention in the discourse (discourse anaphora):

- (1) A man came in. **This man** looked familiar to me.

Level 2 shifts pertain to pragmatic definiteness.

Both types of definiteness can be represented on a scale, with the dividing line cutting through some more ambiguous cases such as, typically, definite associative anaphora (DAA). A model close to the one proposed by Czardybon (2017, p. 26) is given here and an example is provided for each case:⁴

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*)

The ambiguous, borderline cases (underlined in the text) are those of sortal (and relational) nouns made unique through the presence of a clause (typically, an establishing relative clause), different types of DAAs and the so-called “complex individual concepts”, that is nouns (sortal and relational ones) made definite through the presence of a particular class of modifiers such as, typically, ordinal numerals, superlatives and restrictive apposition.⁵ Regarding DAAs, the definiteness of the head noun is triggered by “a hidden link or anchor which has to be introduced earlier” (Löbner, 1998, p. 1). The following is a typical example:

- (2) We arrived in an unknown village. **The church** was located on a hill.

As the scale suggests, three types of DAAs exist: “situational”, “relational” and “part-whole” DAAs. The above-mentioned example (2) belongs to the third type since the church is part of the village that was previously mentioned. The two other remaining types of DAAs will be expanded on later, using corpus examples. The information to the left of this buffer zone belongs to pragmatic definiteness, while the information to the right of the buffer zone is included in semantic definiteness.

⁴ The modifications are based upon our study of the situation in Czech.

⁵ These are called “non-lexical functional nouns” by Ortmann (2014).

2 General background

Even though the grammaticalization process through which languages acquire definite articles from demonstratives remains far from fully understood, authors seem to agree on a couple of points: a good candidate for the process is usually the distal or the neutral form of the demonstrative; once the demonstrative gets involved in the process, it undergoes a stage of bleaching, losing its typically deictic semantics (the contextual anchoring of the referent), its capacity of coding distance (if it was present), as well as part of its phonetic and morphological substance (cf. Carlier – De Mulder, 2010; 2011; Diessel, 1999; Greenberg, 1978; Himmelmann, 1997; Krámský, 1972; Lyons, 1999). Furthermore, it is a crucial fact that emerging definite articles spread from contexts of pragmatic definiteness to those of semantic definiteness and beyond:⁶

From a diachronic viewpoint, definite determiners indeed spread along the scale [...] from deictic and anaphoric uses to semantically-unique uses like associative anaphora, individual nouns (*the sun*) and generic NPs (*the Panda*) (Hofherr – Zribi-Hertz, 2014, p. 9).

Several authors have pointed out that Czech, so far seen as an article-less language, is or might be developing a definite article (cf. Adamec, 1983; Berger, 1993; Mathesius, 1926; Meyerstein, 1972; Orlandini, 1981; Vey, 1946; Ziková, 2017; 2018). In this respect, it has been put forward that the demonstrative *ten*, mostly unstressed and neutral in regard to distance distinctions, appears in contexts excluding genuine demonstratives.⁷ These contexts, pertaining to situational deixis, endophora (anaphora and cataphora) as well as Himmelmann’s (1996) “recognitional” uses, are mostly those of pragmatic definiteness. Concerning anaphoric reference, *ten* can, for instance, be used in spoken Czech in a situation where several referents are present and each of them belongs to a different class of “virtual reference” (Milner, 1976). In languages with definite articles, a proper demonstrative is ousted from such contexts:

- (3) Koupi-l jsem si aut-o a motork-u.
 buy-PST.1SG AUX REFL car-ACC.SG and motorbike-ACC.SG
 (**To**) **aut-o** jezd-i výborně, ale
 ten.N.NOM.SG car-NOM.SG go-PRS.3SG excellently but
 s (**to**) **motork-ou** jsou
 with ten.F.INS.SG motorbike-INS.SG be.PRS.3PL
 pořád problém-y.
 always problem-NOM.PL

⁶ The CTD theory does not account for the functioning of generic nouns.

⁷ The other major definite demonstrative forms in Czech are *tento/tenhle* for a close referent and *tam-ten/onen* for a distant one. A variety of other, often substandard forms appear in informal spoken Czech.

I have bought a car and a motorbike. The car works fine but there are problems all the time with the motorbike (example inspired by Pešek, 2014).

Similarly, *ten* is perfectly acceptable, if not almost obligatory, in a situation where there is only one eligible referent corresponding to the noun. Imagine a father saying to his child who is playing with a ball:

- (4) D-ej mi *ten* *míč!*
give-IMP.2SG me.DAT ten.M.ACC.SG ball.ACC.SG
Give me the ball!

Berger (1993) even concludes that *ten* has lost its capacity to function as a demonstrative in situational deixis altogether, a claim which is also corroborated by *ten* being unstressed most of the time⁸ and by the existence of other, longer forms (especially) in (spoken) Czech – *tenhle*, *tenhleten*, *tadyhleten* etc. –, more often stressed and functioning as “pure” demonstratives. In Löbner’s distinction, all these contexts either belong to pure pragmatic definiteness or to the borderline cases between semantic and pragmatic definiteness.

There are several approaches to the definition of the criteria which have to be fulfilled in order for a demonstrative to be considered a full-fledged definite article. Some authors (cf. Czardybon, 2017; Diessel, 1999; Lyons, 1999), emphasizing the role of discourse anaphora in the grammaticalization process,⁹ argue that the demonstrative has to become obligatory in anaphoric uses.¹⁰ In contrast to this “minimalist” thesis, Krámský (1972) argues that the former demonstrative has to become systematic with generic uses of nouns. This thesis is far too strict since many languages regarded as having a definite article (such as English) do not use the article with plural generic nouns. Still other authors (cf. Himmelmann, 1997; Carlier – De Mulder, 2010; 2011; Löbner, 1985; 2011) stress the importance of those contexts of use which are shared by pragmatic and semantic definiteness. As Carlier and De Mulder (2011, p. 4) put it:

Neither Lyons’s hypothesis nor the hypothesis deriving the definite article from an anaphoric demonstrative referring to a previously mentioned discourse participant do account for this meaning shift from direct reference to indirect reference, which is crucial in the development of the definite article.

The author of the present study fully identifies with this position since it is the only one that is able to account for the spreading of the former demonstrative to contexts from which it is initially banished, i.e. those of semantic definiteness.

⁸ With the exception of Eastern Moravian and Silesian dialects (Jodas, 2010; personal communication), where *ten* can still function as a stressed demonstrative in situational deixis.

⁹ Lyons (1999) also emphasizes the importance of situational deixis in this process.

¹⁰ For a less radical stance see Laury (1997).

In contemporary informal spoken Czech, *ten* does not appear systematically in contexts of semantic definiteness and it is not even obligatory in anaphoric uses¹¹ (cf. Czardybon, 2017; Zíková, 2017; 2018). Concerning pure semantic definiteness, typically, the neutral situation for individual and functional nouns such as *slunce* ('the sun') or *Johnova matka* ('John's mother') is to appear without *ten*. In spite of these facts, it seems logical to assume the grammaticalization process is in progress, without having reached its more advanced stages yet.¹² Besides a significant increase in the use of *ten* with discourse anaphora (Zíková, 2017; 2018), the best evidence of this is the presence of *ten* in contexts which have something to do with semantic definiteness and in which, as a consequence, genuine demonstratives are either clearly excluded or rather difficult to imagine. This is why, for the research presented in this study, the particularly relevant contexts of use are precisely those which are located in the buffer zone of the definiteness scale.

3 Methodology

As the source of empirical data, we used the corpus ORTOFON v1, a balanced corpus of informal spoken Czech. This corpus was created by the Institute of the Czech National Corpus in Prague and issued in 2017.¹³ The basic assumption underlying this study is that the spoken variety of Czech is the one where the grammaticalization process reaches its highest development. This assumption not only relies on our native speaker intuition but also on 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 already taken into account (the spoken corpora made up one fourth of the data). As a direct consequence of this inclusion of informal spoken language, the lemma *ten* got to the position of the 3rd most frequent lemma.

¹¹ Unlike in Upper Sorbian and Upper Silesian (Czardybon, 2017).

¹² The situation is similar in other Slavic languages such as Polish (especially its Upper Silesian dialect) and Upper Sorbian (Czardybon, 2017). In two other Slavic languages, Bulgarian and Macedonian, the demonstrative paradigm has yielded a full-fledged suffixed definite article, cf. Macedonian *Zemjata se vrti okolu Sonceto* ('Earth turns around the Sun'; Galton, 1973).

¹³ The data was collected in the period between 2012 and 2017. The corpus 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). All the examples that follow are cited from this corpus.

The corpus ORTOFON v1 was annotated offline, after being imported into the TXM software, a freely available textometric software developed by the CACTUS research group at the École normale supérieure in Lyon. The following query was used:

```
[pos="PROV.DEM" & lemme!="takov.*"] [pos!="VER|CON|PRE|NOM|PCT"] {0,3}
[pos="N.*" & word!="vole"]14
```

Although it is far from perfect (it certainly failed to capture all uses of the adnominal *ten* contained in the corpus), it turned out to be the best possible query. As its form suggests, measures had to be taken in order 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. Regardless of these measures, the result still contained about 15 per cent of false positives, mostly those where *ten* was used as a pronoun.¹⁵

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* (see Table 1). A brief description of these categories will now be presented. First of all, the distinction was made between several types of deictic reference. This distinction proceeds from systems proposed by several authors (Bühler, 1965/1934; Fraser – Joly, 1979; 1980; Halliday – Hasan, 1976; Himmelmann, 1996; Kleiber, 1994; Mathesius, 1926), but it also reflects specificities encountered in the analyzed material. Three major types of deictic reference were distinguished: deixis proper (situational and temporal), endophora (including anaphora and cataphora), and the “recognitional” uses (Himmelmann, 1996), i.e. those where the identification of the referent relies neither upon the speech situation nor upon information mentioned in the previous discourse, but is based on the shared knowledge of the speaker and the hearer. Within the category of endophora, several sub-types of nominal anaphora were dissociated: faithful (the head noun of the anaphoric expression is the same as that of the antecedent), unfaithful (the head noun of the anaphoric expression is not the same as that of the antecedent), non-coreferential (the anaphoric expression refers to a different entity than the antecedent), temporal anaphora (anaphoric reference to time events), associative anaphora and discourse deixis.¹⁶

¹⁴ The other forms of the paradigm of definite demonstratives were also included into the query so as to compare the relative frequency of all demonstratives to the relative frequency of *ten* alone.

¹⁵ Unfortunately, the tagging of the corpus does not distinguish between adnominal and pronominal uses of demonstratives.

¹⁶ “Discourse deixis is to be understood here as reference to propositions or events” (Himmelmann, 1996, p. 224). Since discourse deixis stands halfway between deixis proper and anaphora, some authors consider it part of the former whereas others see it as part of the latter.

Furthermore, three other phenomena were classified separately, namely “emotional”, “metalinguistic”, and “ambiguous” uses. First, the so called “emotional” or “affective” uses are those where the demonstrative does not participate in the identification of the referent since the referent is already fully identified. Typically, this happens when the demonstrative introduces proper names, generic nouns or nouns with unique reference (*unica*). Generally, this happens every time the noun is purely semantically definite. It might be objected that regarding these cases as “emotional” uses of the demonstrative is controversial, since how can we know that we are not dealing precisely with an advanced stage of the grammaticalization of *ten* into a definite article? However, two major reasons prevent the author of this study from adopting this view. First, it is not very common even for languages with definite articles to use them with proper names.¹⁷ Concerning plural generic nouns, only some languages with articles (such as the Romance languages) use definite articles. Second and most importantly, it seems improbable to assume that Czech – which, as already mentioned, does not even make systematic use of the article-like *ten* with pure pragmatic definiteness – should employ it in contexts of pure semantic definiteness. Indeed, such an assumption would be in direct contradiction to the assumption expressed in the quote from Hofherr and Zribi-Hertz (2014) above, i.e. that the grammaticalization process is basically a gradual one. An analogous position is adopted by Czardybon (2017, p. 75) for Polish:

The first reason for the occurrence of *ten* in cases of semantic uniqueness is when the speaker is emotionally affected. This can be a positive or negative emotion such as happiness or anger. In such an emotional context, *ten* can occur regardless of the concept type since *ten* functions to indicate emotional involvement and not unique reference.

The case of “emotional” uses will be addressed in more detail at the end of this paper, where it will be argued that it may have some importance for the later stages of the grammaticalization process.

The “metalinguistic” uses are those where *ten* is used by a speaker who cannot find the right word or expression. There are phenomena accompanying this type of use, such as, typically, hesitation sounds, repetitions, and the stressing of *ten*.

A certain number of occurrences could not be assigned to a category (the context turned out insufficient in this respect). These were marked as “ambiguous”. In general, the analysis of the occurrences and their assignment to a particular type and/or subtype turned out a rather delicate task. This was partly due to the inherent ambiguity which accompanies the use of the demonstrative in natural language, contrary to artificial examples. But it was also due to the limits of the corpus itself: since the user often lacks a broader context, he or she cannot always know whether a referent

¹⁷ However, such cases do exist: for instance, Modern Greek and some dialects of German.

introduced by a demonstrative was already mentioned before the recording took place, thus being unable to determine whether its identification relies on endophora or a recognitional use. It is similarly impossible to watch the speakers' interaction, which makes it sometimes difficult to determine whether a referent is present in the situation and is thus identified via situational deixis. Distinguishing between cases of DAA and "emotional" uses is another sensitive issue. In the following example, the speakers discuss different approaches to teaching. One of them argues that in order to make students active in class, one has to put them into an unusual situation:

| | | | | | | |
|-----|------------------|---------------|------------------|-----------------------|---------------|--------------------|
| (5) | je | dobr-é | prostě | vždycky | nějak-ým | |
| | be.PRS.3SG | good-N.NOM.SG | MOD | always | some-M.INS.SG | |
| | způsob-em | uvést | do | nestandardn-í | | |
| | manner-INS.SG | put.INF | into | non-standard-F.GEN.SG | | |
| | situac-e | že | se | mus-í | prostě | nějak |
| | situation-GEN.SG | that | REFL | must-PRS-3PL | MOD | somehow |
| | jako | projevit | jako | třeba | nejlepš-í | je |
| | MOD | manifest.INF | MOD | for example | best-N.NOM.SG | be.PRS.3SG |
| | prostě | fakt | jako | aby | se | hýba-l-i |
| | MOD | MOD | MOD | that | REFL | move-PST-3PL.M |
| | po | tě | tříd-ě | | [...] | vsta-l-i |
| | about | ten.F.LOC.SG | classroom-LOC.SG | | [...] | stand-up-PST-3PL.M |
| | z | těch | lavic | jako že jo | | protože |
| | from | ten.F.GEN.PL | desk.GEN.PL | MOD | | because |
| | v | těch | lavic-ích | jsou | | |
| | in | ten.F.LOC.PL | desk-LOC.PL | be.PRS.3PL | | |
| | prostě | fakt | jako | mas-a | ne? | |
| | MOD | really | like | mass-NOM.SG | MOD | |

it's always good to like put them into some non-standard situation so that they like have to express themselves like for example the best way is to make them move about the classroom or to make them stand up from their desk because you see when sitting at the desk they're really pretty much like a mass

The presence of *ten* with the nouns *třída* ('classroom') and *lavice* ('desks') can either be accounted for as an "emotional" use or as DAA. In the first interpretation, the speaker uses *ten* with generic reference in order to mark her emotional involvement with it. In the second interpretation, *ten* marks the referents as definite since they are typical parts of the general and stereotypical situation of teaching at school, including also other elements such as the teacher, the students, and the blackboard. Eventually, the occurrences were considered as "emotional" uses, since the speaker's tone together with the frequent lexical modality bolster this interpretation.¹⁸

¹⁸ In general, having the corpus recordings and full texts at our disposal proved very useful while evaluating the occurrences.

All the occurrences were also examined in terms of the putative grammaticalization. Every time it was assumed *ten* was on its way to become a definite article, it was assigned the GRAM mark. Naturally, this was partly based on our personal assessment and therefore cannot be regarded as completely objective. The description of the main criteria on which the assessment relied will now be provided.

In the case of situational deixis, we followed Berger's (1993) claim that *ten* never acts as a demonstrative in this context and therefore is grammaticalized. If the occurrence is translated into an article language, *ten* will be replaced by a definite article. In the following example, two speakers talk about orchids that one of them grows at home. The orchids are present in the same room as the speakers:

- (6) S1: no jo já to m-á-m proti
 MOD I.NOM that.N.ACC.SG have-PRS-1SG opposite
- východ-u jenže když je
 east-DAT.SG but when be.PRS.3SG
- to tam na **tom**
 that.N.NOM.SG there on ten.M.LOC.SG
- stolečk-u** u **toho** **radiátor-u**
 small table-LOC.SG by ten.M.GEN.SG radiator-GEN.SG
- S2: tak oni ty orchidej-e
 MOD they.NOM ten.F.NOM.PL orchid-NOM.PL
- ne-m-a-jí takov-ý ne ne-m-a-jí
 NEG-have-PRS-3PL such-M.ACC.SG NEG NEG-have-PRS-3PL
- rád-y takov-ej žár v-í-š?
 glad-F.PL such-M.ACC.SG heat.ACC.SG know-PRS-2SG

S1: well (I have put them in such a way that) they're situated towards the east but when they're on the small table by the radiator

S2: well orchids don't like [...] I mean they don't like this much heat you know?

Given the fact that the small table and the radiator are apparently the only representatives of their respective referential classes within the speech situation, article languages will spontaneously use the definite article here; yet the speakers used *ten* in both cases. Coming back to Berger's (1993) claim that *ten* can no longer act as a true demonstrative in situational deixis, the corpus material of this study seems to confirm this since every time the situational referent is contrasted to other referents of the same referential class, a longer and potentially stressed demonstrative form is used. This happens in the following example, where the speaker is talking about a skirt depicted in a photo which she and the interlocutor are looking at:

- (7) no vid-í-š **tahle** **sukn-ě**
 MOD see-PRS-2SG tenhle.F.NOM.SG skirt-NOM.SG
- už mi je teď'ka velk-á
 already me.DAT be.PRS.3SG now big-F.NOM.SG
- see this skirt is already too loose for me now

Since the female speaker is talking about a particular skirt that she (implicitly) contrasts with other skirts, she uses the stressed and longer demonstrative form *tenhle*.¹⁹

Concerning anaphoric occurrences,²⁰ we largely relied on the findings of Laury (2017) and Zíková (2017; 2018). The former argues that the Finnish demonstrative *se* acquired the function of a definite article once it started to mark all identifiable referents and not just prominent referents reappearing in the discourse after a moment of absence (Laury, 2017, p. 147). The latter starts from the premise that if a word is still a typical demonstrative, its occurrence with repetitive mentions should be limited. On the other hand, if it has ceased to be one, it tends to appear more often with repetitive mentions, where it serves as an explicit definiteness marker (Zíková, 2018, p. 110). Thus, when the speakers used *ten* with the second mention of a referent in order to establish it as known or when they used it to reintroduce an important referent after some time had elapsed from its previous mention, *ten* was regarded as a demonstrative. But if *ten* appeared more than twice with the same referent and if this referent was still strongly activated in these subsequent mentions, *ten* was considered close to a definite article and therefore undergoing grammaticalization. In the following example, the referent of *tapety* ('wallpapers') is first introduced by a bare noun. Twelve clauses later, the noun is repeated with *ten*. This second mention establishes the referent as known – *ten* is therefore considered a demonstrative. But only five clauses later, after the referent has been referred to using the pronominal *ten* (*to* and *to všechno*), the same noun is again preceded by *ten*. At this point, not only is the referent established as identifiable, but it is still strongly activated in the hearer's mind, and so *ten* is closer here to a definite article:

- (8) S1: ti soused-i jak jsem
 ten.M.NOM.PL neighbor-NOM.PL that AUX
 ti vo nich vyprávě-l-a
 YOU.DAT about them.LOC tell-PST-3SG.F
 jak tam bydle-l-i u mě
 that there live-PST-3PL.M by me.GEN
 ty NP²¹ jak voni mě-l-i
 ten.M.NOM.PL NP that they.NOM have-PST-3PL.M
 tapet-y tam padesát let
 wallpaper-ACC.PL there fifty years.GEN
 S2: no tak to já je ne*
 mod I.NOM them.ACC NEG

¹⁹ Concerning *ten*, an overwhelming majority of its occurrences in the sample were unstressed. The only exceptions were the “metalinguistic” uses, where the stress can be easily accounted for in terms of prosodic and pragmatic constraints, and one example where *ten* was stressed while used with situational deixis.

²⁰ Except for DAA.

²¹ In ORTOFON v1, “NP” stands for anonymized proper names of persons.

| | | | | | | |
|-----|----------------|--------------|--------------------|------------------|----------------|-------|
| S1: | a | ona | ještě | by-l-a | kuřák | |
| | and | she.NOM | still | be-PST-3SG.F | smoker.NOM.SG | |
| | a | vona | rád-a | děla-l-a | tydlet-y | |
| | and | she.NOM | glad-F.SG | make-PST-3SG.F | these-F.ACC.PL | |
| | panenk-y | | víš | | | |
| | doll-ACC.PL | | MOD | | | |
| S2: | já | v-í-m | | | | |
| | I.NOM | know-PRS-1SG | | | | |
| S1: | ona | skoro | by-l-a | už | na | |
| | she.NOM | almost | be-PST-3SG.F | already | on | |
| | smrt | | nemocn-á | | | |
| | death.ACC.SG | | ill-F.NOM.SG | | | |
| S1: | já | jsem | tam | přiš-l-a | | |
| | I.NOM | AUX | there | come-PST-3SG.F | | |
| | ona | mi | ukazova-l-a | panenk-y | | |
| | she.NOM | me.DAT | show-PST-3SG.F | doll-ACC.PL | | |
| | já | řik-á-m | no | Dáš-o | | |
| | I.NOM | say-PRS-1SG | MOD | Dáša-VOC | | |
| | ale | je | to | na | | |
| | but | be.PRS.3SG | that.N.NOM.SG | on | | |
| | úkor | | se | urazi-l-a | | |
| | expense.ACC.SG | | REFL | offend-PST-3SG.F | | |
| | no | poněvač | a | teď | tam | |
| | MOD | because | and | now | there | |
| | přiše-l | | nov-ej | majitel | | |
| | come-PST.3SG | | new-M.NOM.SG | owner.NOM.SG | | |
| | a | ty | představ | si | | |
| | and | ten.F.ACC.PL | imagine.IMP.2SG | REFL | | |
| | ty | | tapet-y | muse-l | | |
| | ten.F.ACC.PL | | wallpaper-ACC.PL | must-PST.3SG.M | | |
| | strhat.INF | | to | by-l-y | ještě | |
| | tear off | | that.N.NOM.SG | be-PST-3PL.F | still | |
| | takov-ý | | ze | západní-ho | | |
| | such-F.NOM.PL | | from | Western-N.GEN.SG | | |
| | Německ-a | | rozum-i-š | proto | mi | tam |
| | Germany-N.GEN | | understand-PRS-2SG | that's why | me.DAT | there |
| | děla-l-y | | sbíječk-y | to | | |
| | do-PST-3PL.F | | jackhammer-NOM.PL | that.N.NOM.SG | | |
| | všeck-o | | š-l-o | dolů | to | |
| | all-N.NOM.SG | | go-PST-3SG.N | down | that.N.NOM.SG | |
| | by-l-o | | příšern-ý | tak | dlouho | |
| | be-PST-3SG.N | | horrible-N.NOM.SG | so | long | |
| | ty | | tapet-y | tam | | |
| | ten.F.NOM.PL | | wallpaper-NOM.PL | there | | |
| | by-l-y | | že | jo | | |
| | be-PST-3PL.F | | MOD | | | |

- S1: those neighbors I told you about those who lived at my place those NP they had wallpapers there for fifty years
 S2: well I don't kn*
 S1: and on top of that she was a smoker and she liked making these dolls you know
 S2: I know
 S1: she was almost deathly ill I came there she showed me dolls I told her but Dáša this is at the expense she got offended because well and now a new landlord arrived there and imagine he had to tear off those wallpapers plus these were wallpapers from West Germany you see that's why I had jackhammers working around my place it all had to be taken down that was horrible the wallpapers had been there for so long right

Some occurrences of *ten* with non-coreferential anaphora were also analyzed as grammaticalized. For some of these, the term “partially coreferential” rather than “non-coreferential” should be used since the second mention picks up part of the reference of the antecedent:

- (9) Katka m-á na mě-l-a tady
 Katka have-PRS.3SG on have-PST-3SG.F here
 na návštěv-ě dv-a Francouz-e
 on visit-LOC.SG two.M.ACC.PL Frenchman-ACC.PL
 jak by-l-a ve Franci-i tak tam chodi-l-i
 as be-PST-3SG.F in France-LOC.SG so there go-PST-3PL.M
 lézt na nějak-ou stěn-u
 climb-INF on some-F.ACC.SG wall-ACC.SG
 [...]
ten ***jeden*** je student
 ten.M.NOM.SG one.M.NOM.SG be.PRS.3SG student.NOM.SG
 a jeden už prac-uj-e
 and one.M.NOM.SG already work-PRS-3SG
 Katka has [...] she has recently hosted two French guys; when she was in France they used to go climbing there on a climbing wall one (of them) is a student and one already works

Assessing grammaticalization turned out to be the most delicate issue with recognitional uses. Generally speaking, an argument of some value for seeing an occurrence of *ten* as engaged in grammaticalization is its replacement by a definite article once the sentence is translated into an article language such as English or French. However, this criterion is only a very rough and approximate one. One of the reasons why this is so is that speakers of article languages often have a choice between the definite article and the demonstrative when introducing a referent whose identification relies on their shared knowledge. In this respect, *ten* is ambiguous since it may be translated by both forms. In the following example, the speakers are reminiscing about what is, at the same time, depicted in a photo in front of them (yet the referent is identified via the memory of a shared experience):

| | | | | |
|----------|--------------|-------------------|--------------|-------------------|
| (10) S1: | tady | jsou | <i>ty</i> | <i>promenád-y</i> |
| | here | be.PRS.3PL | ten.F.NOM.PL | promenade-NOM.PL |
| | ne | kolem | <i>ty</i> | <i>chodník-y</i> |
| | MOD | around | ten.M.NOM.PL | pavement-NOM.PL |
| S2: | jo | tadyhle | je | |
| | MOD | here | be.PRS.3SG | |
| | <i>ta</i> | <i>promenád-a</i> | no | |
| | ten.F.NOM.SG | promenade-NOM.SG | MOD | |

S1: see here are the/those promenades and the/those pavements around you see

S2: right here is the/that promenade

Both forms are possible in English, depending on the anticipated cost of the activation of the referent for the hearer. This brings us to one of the limits of the present study. Yet an important question is whether the presence of *ten* is felt obligatory from the perspective of a Czech speaker. The referents of *ty promenády* and *ty chodníky* are discourse new but identifiable. This identifiability is most probably a case of pragmatic definiteness. Personally, we tend to see the presence of *ten* here as rather obligatory. If *ten* were absent from this context, the nouns would be interpreted either as referring to semantically definite “permanently established individual concepts” (Gerland – Horn, 2010) or – given their rhematic position in the sentence – as indefinites. In fact, the frequent presence of the weakened form of the demonstrative with discourse new yet pragmatically identifiable referents seems to be a typical feature of languages in which the demonstrative undergoes grammaticalization. This was observed by Laury (1997) in Finnish data. Laury compared two sets of Finnish narratives dating, respectively, from the 1800s and the 1930–40s as to the occurrence of the developing definite article *se*:

[...] in the earlier narratives, no noun phrases with new but identifiable referents had been *se*-marked, but in the data from the 1930s and 40s, a full 60% of such noun phrases are *se*-marked (Laury, 1997, p. 189).

It was therefore concluded that all pragmatically definite referents introduced by *ten* which, at the same time, were identified independently of either the ongoing discourse or the speech situation participated in the grammaticalization process.²²

Whenever it was found that an occurrence belonged to one of the three cases situated between pragmatic and semantic definiteness mentioned above, the presence of *ten* was regarded as grammaticalized. These cases will be presented in more detail in the following section.

²² This class included most of the recognitional uses of *ten*, apart from those where *ten* was preceded by *nějaký* (‘some’) or *takový* (‘such’). The combinations of *nějaký + ten* and *takový + ten* are more problematic since the first element is indefinite (see Uhlířová, 1992).

4 Research results

The query resulted in approx. 20,530 occurrences of adnominal demonstratives. Of these, occurrences of *ten* represented approx. 19,800, a figure which makes up a relative frequency (occurrences per million positions) of app. 16,013. The relative frequency of all demonstratives in adnominal uses was approx. 16,603.

Table 1 below summarizes the number of occurrences of *ten* presumably engaged in the grammaticalization process in proportion to the whole of the 1,000 occurrences of *ten*. It can be seen that the occurrences of *ten* undergoing grammaticalization make up more than half of all the occurrences analyzed (precisely, 53.6%). Most of them appear in endophoric and recognitional uses. Their lower use with deixis is undoubtedly due to two facts: first, deictic uses are in general less represented in the corpus than the two other types and, second, situational deixis is quite often marked by the complex demonstrative forms.

| Type and subtype | Occurrences undergoing grammaticalization | % of all occurrences | Total |
|-------------------------------|---|----------------------|--------------|
| Ambiguous occurrences | | 0.00 | 69 |
| Deixis | 43 | 4.30 | 44 |
| Situational deixis | 43 | 4.30 | 43 |
| Temporal deixis | | 0.00 | 1 |
| Endophora | 194 | 19.40 | 252 |
| Definite associative anaphora | 47 | 4.70 | 47 |
| Faithful anaphora | 60 | 6.00 | 83 |
| Unfaithful anaphora | 5 | 0.50 | 13 |
| Non-coreferential anaphora | 14 | 1.40 | 22 |
| Discourse deixis | | 0.00 | 12 |
| Temporal anaphora | | 0.00 | 6 |
| Cataphora | 68 | 6.80 | 69 |
| Recognitional uses | 299 | 29.90 | 329 |
| Emotional uses | | 0.00 | 292 |
| Metalinguistic uses | | 0.00 | 14 |
| Of all occurrences | 536 | 53.60 | 1.000 |

Table 1: The proportion of the occurrences of *ten* presumably undergoing grammaticalization for each particular type of use in 1,000 randomly chosen occurrences of *ten* (source: ORTOFON v1).

Let us now move on directly to the grammaticalization of *ten* in those endophoric contexts which are situated in the buffer zone between pragmatic and semantic definiteness, that is (1) sortal and relational nouns with establishing relative clauses, (2) the so-called “complex individual concepts”, and (3) DAAs. The first and the second case will be addressed together as they can both be subsumed under the more general term “structural cataphora” (cf. Halliday – Hasan, 1976). Within the sample,

68 occurrences were analyzed as representing this phenomenon (actually, all the cataphoric occurrences of *ten* except for one). These are the proportions of the different modifiers establishing the head noun as having unique reference:

| | |
|---|-----------|
| Superlatives | 4 |
| Ordinal numerals | 18 |
| Other terms | 13 |
| <i>samý/stejný</i> ('same') | 2 |
| <i>další</i> ('next') | 2 |
| <i>budoucí</i> ('future') | 1 |
| <i>předtím</i> ('before'/'previous') | 2 |
| <i>hlavní</i> ('main') | 1 |
| <i>původní</i> ('original') | 2 |
| <i>starý</i> (in the sense of 'former') | 3 |
| Establishing relatives | 28 |
| Complement clauses | 5 |

Table 2: The occurrences of *ten* undergoing grammaticalization due to the presence of a restrictive modifier (cases of “structural cataphora”) in 1,000 randomly chosen occurrences of *ten* (source: ORTOFON v1).

Some examples follow. First, a lady is showing a picture of billboards in Los Angeles to another interlocutor:

- (11) jo a todleto#s jsi vidě-l-a ? jak
 MOD this.N.ACC.SG AUX see-PST-2SG.F how
 v tý na **tý** **nejrušnějš-í**
 in ten.F.LOC.SG on ten.F.LOC.SG busiest-F.LOC.SG
ulic-i v Los Angeles jsou
 street-LOC.SG in Los Angeles.LOC.SG be.PRS.3PL
 takov-ý-dle upoutávk-y
 such-F.NOM.PL-DEICTIC SUFFIX billboard-NOM.PL
 and did you see this? these billboards on the busiest avenue in Los Angeles?

Another example is taken from a conversation about the vicious circle of indebtedness. The noun is made definite through the presence of a relative clause:

- (12) přijd-u znova třeba znova o
 lose-PRS.1SG again MOD again about
 zaměstnán-í znova že jo takže prostě
 job-ACC.SG again MOD so MOD
 jakoby tu a tam něco zaplat-í-m
 MOD now and then something.ACC pay-PRS-1SG
 ale **ten** **dluh** **co** **tam**
 but ten.M.ACC.SG debt.ACC.SG which.NOM there

| | | | |
|--------------|---------------|-----------------|------------|
| <u>by-l</u> | ten | ne-můž-u | platit.INF |
| be-PST.3SG.M | this.M.ACC.SG | NEG-can-PRS.1SG | pay |

then I can lose my job again right so like I'll pay off something now and then but I can't pay the debt which already existed before

The following example is about a character from Jack London's novels who developed a survival mechanism consisting in piling up food at times of abundance. The noun is made definite in Czech through the presence of a complement clause introduced by *že* ('that') ():

| | | | | | |
|------|-----|--------|-----------|-------------|-------|
| (13) | ale | on | samozejmě | m-á | potom |
| | but | he.NOM | of course | has-PRS.3SG | then |

| | | |
|--------------|--------------------|----------------|
| ten | psychick-ej | problém |
| ten.M.ACC.SG | mental-M.ACC.SG | problem.ACC.SG |

| | | | |
|-----------|-------------|-----------|--------------|
| <u>že</u> | <u>jako</u> | <u>si</u> | <u>znova</u> |
| that | MOD | REFL | again |

| | | | |
|---------------------|--------------|-------------|-----------------|
| <u>připrav-uj-e</u> | <u>kdyby</u> | <u>by-l</u> | <u>náhod-ou</u> |
| prepare-PRS-3SG | in case.COND | be-PST.3SG | by chance |

| | |
|------------------|---------------|
| <u>dalš-í</u> | <u>hlad</u> |
| another-M.NOM.SG | famine.NOM.SG |

but then he has the mental problem so he stocks up on food again just in case another famine came

The occurrences of *ten* with DAA are also relatively numerous in the sample: 47 of them were identified.²³ As already stated above, the present study subscribes to Czardybon's terminology, distinguishing between "part-whole", "relational" and "situational" DAA (Czardybon, 2017). It is worth noting that the occurrences in the sample belong either to the situational or to the relational subtype, while none belong to the part-whole subtype. The following two examples show a situational and a relational DAA, respectively. The first is about a room full of cobwebs and the spiders that have woven them:

| | | | | | | |
|------|-----|----------------------|-------|-------------------|---------------|-------|
| (14) | S1: | nahoře | nad | tím | by-l | |
| | | upstairs | above | that.N.INS.SG | be-PST.3SG | |
| | | by-l-y | tak | obrovitánsk-ý | pavučin-y | [...] |
| | | be-PST-3PL.F | so | gigantic-F.NOM.PL | cobweb-NOM.PL | [...] |
| | S2: | jako že | se | tam | rok-y | |
| | | MOD | REFL | there | years-ACC.PL | |
| | | ne-vysáva-l-o | | | jo? | |
| | | NEG-vacuum-PST-3SG.N | | | MOD | |

²³ However, it often turned out rather difficult to decide whether an occurrence was really a case of DAA rather than an emotional use or a recognitional use.

S1: ne on ten týpek
 no he.NOM ten.M.NOM.SG guy.NOM.SG
 říka-l že *ti* *pavou-ci*
 say-PST.3SG that ten.M.NOM.PL spider-NOM.PL
 to uděla-l-i za hodin-u
 that.N.ACC.SG make-PST-3PL.M in hour-ACC.SG
 takov-ou obrovsk-ou pavučin-u
 such-F.ACC.SG huge-F.ACC.SG cobweb-F.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

In the following example, the speaker compares bachelor's degree programmes to studies at a vocational school in the Czech Republic in terms of the demands they place on the students:

- (15) no tak t* já jsem jí říka-l-a
 MOD I.NOM AUX her.DAT tell-PST-1SG.F
 že to už m-á úplně
 that that.N.ACC.SG already have-PRS.3SG completely
 to sam-é když
 ten.N.ACC.SG same-N.ACC.SG if
 pů-jd-e studovat bakalář-e
 FUT-go-3SG study-INF bachelor's-ACC.SG
 a bude to mít-INF takov-ou
 and AUX that.N.NOM.SG have have such-F.ACC.SG
 jakoby větš-í váh-u než
 MOD bigger-F.ACC.SG weight-ACC.SG than
 ta vošk-a přitom [...] *ta*
 ten.F.NOM.SG vocational school-NOM.SG yet [...] ten.F.NOM.SG
náročnost je stejn-á
 difficulty.NOM.SG be.PRS.3SG same-F.NOM.SG
 si mysl-í-m jakoby
 REFL 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

While the head of a part-whole or a relational DAA is an inherently relational noun, the head of a situational DAA is a non-relational noun shifted into a relational concept (Czardybon, 2017, pp. 67–71): the noun *náročnost* ('difficulty') is an inherently relational noun since it calls for an argument (*the difficulty of something*); *spiders* are not a relational noun, yet the noun is shifted into a relational concept since the

speaker talks about the spiders present in the particular situation (“the spiders of the particular cobwebs”). In terms of the progress of grammaticalization, Czardybon (2017, pp. 65–72, 79–91) uses data from several Slavic languages to show that the part-whole subtype is situated closest to the pole of semantic definiteness while the situational subtype is furthest from it. Therefore, it is not surprising that the situational type should be the most abundant type in the sample (33 occurrences).²⁴ But contrary to Czardybon’s (2017, p. 90) conclusion, according to which the relational type is impossible in spoken Czech, our sample still contains 14 occurrences of it. In any case, the presence of *ten* with DAA constitutes the most eloquent proof of its ongoing grammaticalization into a definite article since, as it is well known, genuine demonstratives are ousted from this context of use. Should a form such as *tihle* (‘these’) replace *ten* in the spider example, the reference of the NP *tihle pavouci* (‘these spiders’) could not be interpreted as a case of DAA.

Let us now briefly finish with the issue of the so-called “emotional” uses of the adnominal demonstrative *ten*. These are known to be very extensive in colloquial Czech (cf. Mathesius, 1926; Adamec, 1983; Berger, 1993; Štícha, 1999; Šimík, 2015). In the corpus sample, 284 were identified (almost 30%). Here are three examples:

- (16) to ne to ne v Německ-u je-d-u určitě
 MOD in Germany-LOC.SG drive-PRS-1SG certainly
 cel-ou dob-u po dálnic-i tak
 whole-F.ACC.SG time-ACC.SG on motorway-LOC.SG so
 to není že jo problém
 that.N.NOM.SG be.PRS.3SG.NEG MOD problem.NOM.SG
 do *toho* *Hannover-u*
 to ten.M.GEN.SG Hannover-GEN.SG

in Germany I’ll be driving on the motorway all the time so getting to Hannover won’t be a problem

- (17) ani ne-chtě-l před pr* před
 even NEG-want-PST.3SG before before
 Vánoc-ama nastúpit do *té*
 Christmas-INS.PLTANTUM start-INF into ten.F.GEN.SG
nemocnic-e že si chc-e ještě
 hospital-GEN.SG that REFL want-PRS.3SG still
 posledn-í Vánoc-e užít doma
 last-F.ACC.SG Christmas-ACC.PLTANTUM enjoy-INF at home

he didn’t even want to go to hospital before Christmas saying he wanted to enjoy his last Christmas at home

²⁴ However, the classification sometimes turned out very delicate and controversial here, too.

- (18) nebo jednou mi tak přeběh-l-o děck-o
 or once me.DAT MOD CROSS-PST-3SG.N kid-NOM.SG
 na posledn-i chvíl-i *ta* *jeho*
 at last-F.ACC.SG moment-ACC.SG ten.F.NOM.SG his.NOM
matk-a se vůbec ne-stara-l-a
 mother-NOM.SG REFL at all NEG-care-PST-3SG.F
 or once a kid crossed the street in front of me at the last moment [...] that mother of his
 didn't care at all

The feature these three examples have in common is that *ten* introduces into the discourse a semantically definite noun: in (16), it is a toponym referring to a unique place on Earth. Next in (17), the noun *nemocnice* ('hospital') stands for an individual concept within the pragmatic set shared by the speaker and the hearer. In the last example (18), *ten* is used with a functional noun whose argument is saturated by the adnominal possessive (*jeho*). Since the presence of this possessive suffices to establish the reference of the noun as unique, the procedural semantics of *ten* – whose normal function is otherwise to identify the referent by resorting to the context (cf. Corblin, 1987; Kleiber, 1983) – seems redundant here. At the same time, there seems to be little doubt that in such cases, the role of *ten* has so far nothing to do with pure definiteness marking, which is the role of definite articles. It remains a full-fledged demonstrative signaling, on the part of the speaker, a certain type of emotional involvement with the referent. More precisely, it is the contextual grounding of the referent that allows for the rendering of this emotional involvement.

However, the borderline between the “emotional” and the article-like uses of *ten* may sometimes become blurred. This may be the reason why some authors assert that the (main) function of *ten* with complex individual concepts is to indicate emotional affectedness (cf. Czardybon, 2017; Mathesius, 1926; Zubatý, 1916). This conclusion might sound a bit precipitate. Rather, it should be said that the primary function of *ten* in these contexts is to mark definiteness, while the emotional value can be accessorially added to it. In any case, this assertion seems to suggest that the two phenomena may come into interaction and probably even merge to some extent in some contexts where *ten* is used. When this happens, it might be the case that the emotional value – due to its extensive presence in informal spoken language – acts as an encouraging factor in the grammaticalization process itself, also primarily present in this language variety. Later, when this process reaches more advanced stages, “emotional” uses might also encourage the spreading of *ten* to contexts of purely semantic definiteness since, in these contexts, *ten* can already be present, although with a different original function. This hypothesis would, after all, be in line with Traugott's (1995) approach, placing emphasis on the role that subjectivity and subjectification play in the grammaticalization process.

5 Conclusion

This study has attempted to provide some evidence for the hypothesis according to which Czech is developing a definite article out of the distance-neutral demonstrative *ten* in adnominal positions. Informal spoken Czech was used since this variety is the one most relevant to the grammaticalization process. A sample consisting of 1,000 randomly chosen occurrences showed that more than 50% of them may be seen as affected by the phenomenon. From the point of view of the CTD, considered a decisive theoretical framework in this study, the contexts where *ten* is used either pertain to pure pragmatic definiteness or to cases situated halfway between pragmatic and semantic definiteness. The latter represent a particularly strong piece of evidence for the grammaticalization hypothesis, since these cases are more or less reluctant to accept genuine demonstratives. Concerning the presence of *ten* with pure semantic definites, this case is regarded, in this paper, as one situated outside of the grammaticalization process and pertaining to “emotional” uses. The demonstrative still functions as a pure deictic, anchoring the referent explicitly in the context, although it does not play any role in its identification. However, the claim has been made that these two phenomena might get more interwoven in the future. At that stage, the “emotional” uses might accelerate the grammaticalization process. Yet this claim remains speculative, and, therefore, more research needs to be done here.

REFERENCES

- ADAMEC, Přemysl (1983): České zájmeno *ten* a jeho ruské ekvivalenty. In: Vladimír Hrabě – Aleksandra Grigor'jevna Širokovová (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.
- BÜHLER, Karl (1965/1934): *Sprachtheorie: Die Darstellungsfunktion der Sprache* [2., unveränd. Auflage]. Stuttgart: G. Fischer.
- CARLIER, Anne – DE MULDER, Walter (2010): The emergence of the definite article in Late Latin: “*ille*” in competition with “*ipse*”. In: Hanne Cuykens – Kristin Davidse – Lieven Van de Lanotte (eds.), *Subjectification, Intersubjectification and Grammaticalization*. The Hague: Mouton De Gruyter, pp. 241–275.
- 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.
- CORBLIN, Francis (1987): *Indéfini, défini et démonstratif: Constructions linguistiques de la référence*. Genève – Paris: Droz.
- CZARDYBON, Adrian (2017): *Definiteness in a Language without Articles – A Study on Polish*. Düsseldorf: Düsseldorf University Press.

- 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.
- DIESSEL, Holger (1999): *Demonstratives: Form, Function and Grammaticalization*. Amsterdam – Philadelphia, PA: John Benjamins.
- FRASER, Thomas – JOLY, André (1979): Le système de la deixis: esquisse d'une théorie d'expression en anglais. *Modèles linguistiques*, 1(2), pp. 97–157.
- FRASER, Thomas – JOLY, André (1980): Le système de la deixis: endophore et cohésion discursive en anglais. *Modèles linguistiques*, 2(2), pp. 22–51.
- GALTON, Herbert (1973): The function of the definite article in some Indo-European languages: the grammatical category of determinacy. *Linguistics*, 107, pp. 5–13.
- GERLAND, Doris – HORN, Christian (2010): Referential properties of nouns across languages. In: Y.-S. Kang – J.-Y. Yoon – J. Hong – J.-S. Wu – S. Rhee – K.-A. Kim – D.-H. Choi – K.-H. Kim – H.-K. Kang (eds.), *Universal Grammar and Individual Languages: Proceedings of SICOL 2010*. Seoul: University of Korea, pp. 133–150.
- GREENBERG, Joseph H. (1978): How does a language acquire gender markers? In: Joseph H. Greenberg (ed.), *Universals of Human Language: Volume 3: Word Structure*. Stanford, CA: Stanford University Press, pp. 47–82.
- HALLIDAY, Michael A. K. – HASAN, Ruqaiya (1976): *Cohesion in English*. London: Longman.
- HEIDEN, Serge (2010): The TXM platform: building open-source textual analysis software compatible with the TEI encoding scheme. In: Ryo Otaguro – 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 use. In: Barbara A. Fox (ed.), *Studies in Anaphora*. Amsterdam – Philadelphia, PA: John Benjamins.
- HIMMELMANN, Nikolaus P. (1997): *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Berlin: De Gruyter.
- HOFHERR, Patricia Cabredo – ZRIBI-HERTZ, Anne (eds.) (2014): *Crosslinguistic Studies on Noun Phrase Structure and Reference*. Leiden: Brill.
- JODAS, Josef (2010): Regionální diference v užívání ukazovacích zájmen. *Bohemica Olomucensia*, 2(3), pp. 88–93.
- KLEIBER, Georges (1983): Les démonstratifs (dé)montrent-ils? Sur le sens référentiel des adjectifs et pronoms démonstratifs. *Le français moderne*, 51, pp. 99–117.
- KLEIBER, Georges (1994): *Anaphores et pronoms*. Louvain-la-Neuve: Duculot.
- KOPŘIVOVÁ, Marie – KOMRSKOVÁ, Zuzana – LUKEŠ, David – POUKAROVÁ, Petra – ŠKARPOVÁ, Marie (2017): *ORTOFON: Korpus neformální mluvené češtiny s víceúrovňovým přepisem* [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 16. 5. 2020. <www.korpus.cz>.
- KRÁMSKÝ, Jiří (1972): *The Article and the Concept of Definiteness in Language*. The Hague: Mouton.
- LAURY, Ritva (1997): *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish*. Amsterdam – Philadelphia, PA: John Benjamins.
- LÖBNER, Sebastian (1985): Definites. *Journal of Semantics*, 4(4), pp. 279–326.

- 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.
- LYONS, Christopher (1999): *Definiteness*. Cambridge: Cambridge University Press.
- MATHESIUS, 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.
- MILNER, Jean-Claude (1976): Réflexions sur la référence. *Langue française*, 30, pp. 63–73.
- ORLANDINI, Anna (1981): Wesen und Entwicklung des Artikels vom Lateinischen zu den Romanischen Sprachen. *Indogermanische Forschungen*, 86(1), pp. 223–247.
- 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.
- PEŠEK, Ondřej (2014): Nominální anafora a determinace – kontrastivní analýza francouzských a českých systémových možností. *Časopis pro moderní filologii*, 96(2), pp. 147–164.
- ŠÍMÍK, Radek (2015): On pragmatic demonstratives: the case of pragmatic discourse anaphora in Czech. *Proceedings of Sinn und Bedeutung*, 20, pp. 640–657.
- ŠTÍCHA, František (1999): K deikticko-anaforickým funkcím lexému ten. *Slovo a slovesnost*, 60(2), pp. 123–135.
- TRAUGOTT, Elizabeth Closs (1995): Subjectification and grammaticalisation. In: Dieter Stein – Susan Wright (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge: Cambridge University Press, pp. 31–54.
- UHLÍŘOVÁ, Ludmila (1992): Ten nějaký // nějaký ten a případy podobné. *Naše řeč*, 75(5), pp. 247–254.
- VEY, Marc (1946): *Morphologie du tchèque parlé*. Paris: C. Klincksieck.
- ZÍKOVÁ, Magdalena (2017): *Gramatikalizační potenciál anaforické funkce lexému ten v mluvených narativech* [unpublished doctoral thesis]. Praha: Filozofická fakulta UK.
- ZÍKOVÁ, Magdalena (2018): Anaforická funkce výrazu ten v gramatikalizační perspektivě. *Studie z aplikované lingvistiky / Studies in Applied Linguistics*, 9(special issue), pp. 106–123.
- ZUBATÝ, Josef (1916): Ten. *Naše řeč*, 1(10), pp. 289–294.

Institut d'Histoire des Représentations et des Idées dans les Modernités,
 École normale supérieure de Lyon
 15 parvis René Descartes, BP 7000, 69342 Lyon Cedex 07, France
 jan.dvorak@ens-lyon.fr