

Anett Árvay*, Katalin Nagy C., Tibor Szécsényi and Enikő Németh T.

The analysis of utterances with imperative forms in Hungarian health-related fake news

Felszólító alakokat tartalmazó megnyilatkozások elemzése egészségügyi témájú álhírekben

<https://doi.org/10.1515/lingvan-2024-0065>

Received April 17, 2024; accepted January 15, 2025; published online February 24, 2025

Abstract: This paper aims to contribute to the linguistic detecting of disinformation by providing a corpus-based form-to-function study of Hungarian health-related fake news. It starts from the hypothesis that there is a difference between fake and real news regarding the use of directives as a potential tool of exerting pressure on readers. The most direct and strongest strategy of performing directives in Hungarian is the use of a verb with an imperative suffix. However, the same suffix can also appear without any directive function in subjunctive structures. In order to determine which function utterances containing verbs with an imperative ending have, a manual pragmatic annotation was carried out in our MedCollect corpus. Three major groups of utterances were distinguished: (i) utterances with directive function, (ii) utterances with discourse function, and (iii) utterances without any directive or discourse function. The results support our hypothesis that fake news contains a significantly higher number of directives performed using this direct strategy due to a higher motivation of placing pressure on readers.

Keywords: corpus-based method; directives; imperative form; health-related fake news; Hungarian

Absztrakt: Tanulmányunk a dezinformáció nyelvi felismeréséhez kíván hozzájárulni magyar nyelvű egészségügyi álhírek, „formától a funkcióig” korpuszalapú módszerrel történő vizsgálatával. Abból a hipotézisből indulunk ki, hogy az álhírek és a valódi hírek között különbség van a direktíváknak mint az olvasókra gyakorolt nyomásgyakorlás potenciális eszközeinek használatában. A magyar nyelvben a direktívák végrehajtásának legközvetlenebb és legerősebb stratégiája a felszólító módú igék használata. Ugyanez az igealak azonban direktív funkció nélkül is megjelenhet kötőmódú szerkezetekben. Annak érdekében, hogy meghatározzuk, milyen funkciókkal rendelkeznek a felszólító módú igéket tartalmazó megnyilatkozások, pragmatikai kézi annotációt végeztünk a MedCollect korpuszunkban. A megnyilatkozások három fő csoportját különböztettük meg, (i) a direktív funkcióval rendelkező, (ii) a diskurzusfunkcióval rendelkező és (iii) a direktív és diskurzusfunkció nélküli megnyilatkozásokat. Az eredmények alátámasztották hipotézisünket, azaz, hogy az álhírek szignifikánsan nagyobb számban tartalmaznak közvetlen direktívákat, ami az olvasókra való nyomásgyakorlás magasabb motivációjának tudható be.

Kulcsszavak: korpuszalapú módszer; direktívák; felszólító forma; egészségügyi álhír; magyar

All the authors are members of MTA-SZTE-DE Research Group for Theoretical Linguistics and Informatics.

***Corresponding author: Anett Árvay**, Hungarian Studies Programs, University of Szeged, Egyetem u. 2, Szeged 6722, Hungary, E-mail: anett.arvay@hstud.u-szeged.hu. <https://orcid.org/0000-0002-0215-421X>

Katalin Nagy C., Tibor Szécsényi and Enikő Németh T., Department of General Linguistics, University of Szeged, Szeged, Hungary, E-mail: nagykati@hist.u-szeged.hu (K. Nagy C.), szecsényi.tibor@szte.hu (T. Szécsényi), nemethen@hung.u-szeged.hu (E. Németh T.). <https://orcid.org/0000-0002-1106-7644> (K. Nagy C.). <https://orcid.org/0000-0002-7679-8317> (T. Szécsényi). <https://orcid.org/0000-0001-8559-7229> (E. Németh T.)

1 Introduction

In recent years, the amount of health-related fake news has increased, especially during the COVID-19 pandemic, causing serious harm. Detecting disinformation has become crucial worldwide (United Nations 2023), and not only in the field of healthcare. Fact checking, natural language processing (NLP)-based automatic fake news detection, and linguistic analysis are three potential ways of identifying fake news (see, e.g., <https://library.csi.cuny.edu/misinformation/fact-checking-websites>; Grieve and Woodfield 2023; Scott 2021; Sousa-Silva 2022; Zhou et al. 2019). The first method is very limited and expensive, because it employs the human resources of fact checkers. NLP-based automatic fake news detection algorithms mainly operate through the analysis of the content of texts and certain linguistic features of the news, such as the vocabulary. They can be suitable to identify deception and manipulation in fake news. However, some fake news relies more on partisanship and identity politics than on dishonesty, making it difficult for NLP systems to detect (Sousa-Silva 2022). Since these kinds of fake information are expressed through various linguistic elements, linguistic analysis – including forensic linguistic analysis – seems to be a fruitful method for detecting fake news (Sousa-Silva 2022).

The present paper aims to contribute to the internationally growing field of linguistic analysis of fake news (see, e.g., Grieve and Woodfield 2023; Scott 2021) by investigating linguistic features of Hungarian health-related fake news. The underlying assumption of our project is that there are significant differences between the language use of health-related fake news and real news. One of these differences is the use of directives as a potential tool for putting pressure on readers. Directives are speech acts which the speaker performs in an attempt to get their conversational partner to do something (Searle 1979: 17). Our aim is to identify the utterances that perform directive speech acts in health-related fake and real news. We are proposing the hypothesis that fake news contains significantly more directives than real news due to having a stronger motivation to exert pressure on the reader. We test this hypothesis by analyzing items in the MedCollect corpus of Hungarian health-related articles.

In Hungarian, directives are performed by various strategies. The most direct and strongest one is the use of utterances with direct imperatives, that is, by using verbs with the imperative suffix *-j* and its allomorphs. However, in Hungarian, the use of verbs with the imperative ending is not the only way of performing directives, and directives can be performed by several other strategies as well. In addition, although verbs with an imperative suffix appearing in a main clause generally indicate that the utterance is a directive, they can also appear without any directive function in subordinate clauses in subjunctive structures.¹ Thus, in Hungarian there is no one-to-one correspondence between the occurrences of verbs with the imperative suffix and the directive function of the utterances.

In our research, we applied a “corpus-based form-to-function” pragmatics method (Aijmer 2018). First, by using the morphological analyzer of the Hungarian digital language processing system *e-magyar* (<https://e-magyar.hu/en>),² we automatically searched for the verbs with imperative endings in a part of our MedCollect corpus that was separated out for the analysis of the occurrences of directives. Second, we manually annotated those utterances that contained verbs with imperative endings that had been revealed by the morphological analyzer. We aimed to determine what functions the utterances containing verbs with the imperative suffix have in both fake and real news. Finally, we statistically evaluated the result of the pragmatic annotation.

The organization of our paper is as follows. In Section 2, we describe Hungarian utterances containing verbs with the imperative suffix. Then, in Section 3, we characterize our MedCollect corpus. Section 4 is dedicated to the pragmatic annotation of utterances with directive functions. Section 5 provides the results of the statistical analysis, followed by our conclusions in Section 6.

¹ Subjunctive mood is alternatively labeled as conjunctive mood in the descriptions of some languages.

² *E-magyar* is a toolchain that provides the basic tools for automated processing of Hungarian in an integrated and standardized framework.

2 Verbs with the imperative suffix *-j* in Hungarian

As indicated in Section 1, we hypothesize that fake news contains significantly more directives than real news. Directives are performed in the Hungarian language using various strategies. The most direct and forceful one is the use of verbs with the imperative suffix *-j* and its allomorphs. For a quantitative analysis, one might expect it to be simple to count and compare the number of imperatives in both the fake and the real news texts in the MedCollect corpus. However, this seemingly feasible task was complicated by the fact that automatic identification and counting of the verbs with imperative suffixes would have distorted the data, since in Hungarian these can also occur without any directive function.

Traditional, morphology-based Hungarian descriptive grammars cannot explain this issue – they distinguish only declarative, imperative and conditional moods (e.g., Bencédy et al. 1976; Keszler 2000; Kugler 2000; Simonyi 1895; Tompa 1962) – while syntax-based approaches (e.g., É. Kiss 1998; Hegedűs 2004; H. Varga 2020; Prileszky 1974; Péteri 2012; Szarka 2000; Tóth 2003, 2014) argue for a fourth category, labeling it the subjunctive mood, known in several Indo-European languages. They also state that Hungarian word order indicates the mood, as illustrated in sentences (1) and (2). In (1), the preverb *vissza* ‘back’ precedes the verb; this is the word order of declarative sentences. This is labeled as the subjunctive mood, and it has no directive function; therefore, it does not put any pressure on readers to execute any action. In (2), the preverb *vissza* ‘back’ stands after the verb and indicates the directive function of the utterance; this is called the imperative mood.³

- (1) Function: statement

A könyvtár-ba ment, hogy vissza-vi-gye a könyv-et.
 the library-ILL went that PVB-carry-IMP.DEF.3SG the book-ACC
 ‘S/he went to the library to return the book.’

- (2) Function: directive/performative directive

Kér-em, hogy a hét-en vi-dd vissza a könyv-et a könyvtár-ba!
 please-DEF.1SG that the week-SUP carry-IMP.DEF.2SG PVB the book-ACC the library-ILL
 ‘I ask you to return the book to the library this week.’

Recently, this distinction in terms of word order was challenged by Dömötör and Varga (2014). They have found a few rare cases where discourse particles could cause imperative mood with subjunctive word order (i.e., preverb + verb). Due to the difficulty of separating imperative and subjunctive moods, then, we leave aside the issue of categorization in this respect and concentrate only on the function of utterances. From here, then, all verbs with an imperative suffix will be referred to as imperative verb form, regardless of function.

The fact that there is no one-to-one correspondence between the presence of the imperative suffix and the imperative function in Hungarian forced us to interpret every single utterance in our MedCollect corpus that contains a verb with an imperative suffix, and to check whether the directive function was present or not.

Let us examine what kind of functions can be assigned to the imperative forms and whether these put pressure on readers or not, by examining a set of examples with different functions. Note that the *-j* suffix alters due to phonological rules and can change into *-d*, *-s*, *-gy*, and other similar forms (Kiefer 2018).

- (3) Function: directive/command

Ülj-ön le!
 sit-IMP-INDEF.3SG PVB
 ‘Sit down!’ (formal)

³ Note that the abbreviation *IMP* is used for indicating both the subjunctive and the imperative mood, since in the remainder of the paper we do not distinguish between these two categories. Examples follow the Leipzig Glossing Rules. Abbreviations used: 1/2/3 first/second/third person; ACC accusative case; ALL allative case; DAT dative case; DEF definite conjugation of transitive verb; DEL delative case; ELA elative case; INDEF indefinite conjugation of transitive verb; ILL illative case; IMP imperative or subjunctive mood; INE inessive case; INF infinitive; INS instrumental case; PAST past tense; PL plural; POSS possessive; POT potential; PVB preverb; SG singular; SUB sublativ case; SUP superessive case.

- (4) Function: directive/modal declarative directive
Meg kell hogy ír-j-ad a lecké-t!
 PVB must that write-IMP-DEF.2SG the homework-ACC
 ‘You have to do the homework.’
- (5) Function: directive/impersonal directive
Fontos, hogy az ember alud-j-on egy jó-t.
 important that the people sleep-IMP-INDEF.3SG a good-ACC
 ‘It is important to get a good sleep.’
- (6) Function: asking for advice
Hova ül-j-ek?
 where.to sit-IMP-INDEF.1SG
 ‘Where shall I sit?’
- (7) Function: offer
Ki-nyis-s-am az ablak-ot?
 PVB-open-IMP-DEF.1SG the window-ACC
 ‘Shall I open the window?’
- (8) Function: statement
Éva ar-ra vágy-ik, hogy le-gy-en egy saját csokibolt-ja.
 Éva that-SUB desire-INDEF.3SG that be-IMP-INDEF.3SG an own chocolate.shop-POSS.3SG
 ‘Eva wishes to have her own chocolate shop.’
- (9) Function: statement
Elég nagy ah-hoz, hogy egyedül men-j-en iskolá-ba.
 enough big that-ALL that alone go-IMP-INDEF.3SG school-ILL
 ‘She is big enough to go to school on his/her own.’

In (2)–(5), the imperative verb forms express various types of directive functions, either in a simple sentence like in (3) or in a subordinate clause like in (2), (4), and (5). The types of directive functions in (2)–(5) were categorized according to Vincent et al. (2023). In (2), the verbal structure of the main clause *kérem, hogy* ‘I ask to’, which is labeled as performative directive, requires any verb in the subordinate clause to be in imperative form in Hungarian. It is worth emphasizing that if the main clause with a performative directive verb was deleted, the subordinate clause would turn into a direct directive that is a command just like the example in (3). Example (4) has a modal verb *kell* ‘must’ which requires an imperative verb form. The directive can be categorized as modal declarative directive. Example (5) is an impersonal directive, where the main clause contains the adjective *fontos* ‘important’, which suggests general, rather impersonal advice. Although examples (2)–(5) were all categorized as directives, their forcefulness is not the same. Among them the utterance in (5) implies the softest directive caused by the word *fontos* ‘important’, which decreases the pressure on the reader, whereas the utterance in (3) with the pure imperative has the strongest impact on the reader due to the lack of mitigation. We can conclude that the forcefulness of the directive function can be considered as a scalar notion, and this will be further explored in our MedCollect corpus during the next phase of our research. When it comes to examples (1) and (6)–(9), the imperative verb forms do not have any directive function, but rather they manifest other speech acts, such as asking for advice in (6), offering in (7), or simply asserting something in (1), (8), and (9).

3 MedCollect corpus

The MedCollect corpus contains publicly available online health-related fake and real news from the period between 2007 and 2023. Our goal in building the MedCollect corpus was to enable both qualitative and quantitative

analysis and to test various hypotheses on the linguistic differences between fake and real news. The fake news originates from websites flagged as fake news sites based on fact checking and listed by reliable websites (Urban Legends, <https://www.urbanlegends.hu/tag/atveros-oldalak-listaja>; HVG, https://hvg.hu/tudomany/20150119_atveros_weboldalok). Health-related fake news was manually selected from fake news sites based on their topics, such as epidemics, alternative medications and treatments, and medical devices. Health-related real news, on the other hand, was chosen randomly and was downloaded from reputable health-related websites recommended by our healthcare consultants.

Following the initial selection phase, the members of the research team checked whether the texts labeled as fake news had been categorized correctly. In problematic cases, we asked for the assistance of healthcare experts for classification. Although only the text versions were used during the annotation process, the URLs and PDF files created from the websites are also included in the MedCollect corpus.

The corpus currently contains 1,480 texts (825,467 tokens in total, comprising words, punctuation marks, and symbols), consisting of 675 fake (410,879 tokens) and 805 real news items (414,588 tokens). From the MedCollect corpus, we separated out a smaller subcorpus that underwent a manual pragmatic annotation in order that relevant linguistic features can be marked. The manually annotated subcorpus contains 699 texts (366,832 tokens), including 284 fake (162,391 tokens) and 415 real news (204,441 tokens).

During the manual annotation process, various functions of imperative forms were identified in the texts, which were analyzed by three annotators using the WebAnno annotation tool (Eckart de Castilho et al. 2016). First, the annotation was carried out independently by each annotator with the help of comprehensive annotation guidelines. Second, the annotations were summarized by a curator who gave preference to the majority decision in case of nonuniform labeling. Third, the problematic cases were discussed by the research group. The annotator agreement⁴ was measured regularly.

4 Pragmatic annotation of utterances with imperative forms

The aim of the pragmatic tagging was to find the underlying communicative function of utterances containing verbs with imperative suffixes. During the annotation, the function of each utterance was labeled from the point of view of whether or not it put pressure on the reader to do something at the moment of reading. In addition to marking those utterances with directive function as potential tools for exerting pressure on readers, our other aim was to create theoretically motivated categories. During the pilot annotations, utterances without any directive function where the imperative verb form appeared for purely grammatical reasons were separated out. Then a fine-tuned categorization of the utterances with an imperative verb form that could be associated with a directive was carried out. The latter group's annotation labels were created using the criterion of how pertinent the directives were to the reader and whether or not it prompted them to act.

We categorized the utterances into three main groups: (i) utterances which have a directive function towards the reader at the time of the reading the text, (ii) utterances which fulfill a discourse function, and (iii) utterances which do not have directive or discourse functions. Let us discuss each category.

4.1 Utterances with directive function

The tags *saját hangú* 'own voice' or *közvetített* 'mediated' were given to utterances with directive function, that is, those that are used to put some kind of pressure on the reader at the time of reading. This pressure can come directly from the author who tries to get the reader to do something "with their own voice", or it can be mediated, when the author of the text transmits the directive of someone else. The former cases were labeled as *saját hangú* 'own voice' and the latter ones as *közvetített* 'mediated'. Additional tags were added to indicate each directive's source (single or multiple speakers) and addressee (single or multiple listeners, and also inclusive [1PL] forms).

⁴ For more detailed information, see Szécsényi et al. (2024).

In fake news texts, directives originating from the author frequently occur. Consider (10):

- (10) *Talál-d meg újra az egészség-et,*
 discover-IMP-DEF.2SG PVB again the health-ACC
ami-ben nagymamá-i-nk él-t-ek: a magyar
 which-INE grandmother-POSS.PL-1PL live-PAST-3PL the Hungarian
hagyományos étrend GYÓGYÍTÓ erejű!
 traditional diet healing powered
 ‘Rediscover the health that our grandmothers lived in: the Hungarian traditional diet has HEALING power!’ (MCCfn027/1)

Directives tagged as “mediated” describe or quote a directive utterance performed by another person. In these cases, the author does not distance themselves from the imperative content and mediates the directive force to the readers without dissent, as in (11) and (12).

- (11) *Petz professzor csak ar-ra szeretné figyelmeztet-ni*
 Petz professor only that-SUB would.like-INDEF.3SG warn-INF
a rácsos technológia leendő felhasználó-i-t, hogy
 the grided technology future user-POSS.3PL-ACC that
figyel-j-enek oda ar-ra, nehogy véletlenül
 pay.attention-IMP-INDEF.3PL PVB that-SUB lest accidentally
hamisítvány-t ve-gy-enek meg az Internet-en.
 fake-ACC buy-IMP-INDEF.3PL PVB the internet-SUP
 ‘Professor Petz just wants to warn future users of the grid technology to be careful not to accidentally buy a fake one on the internet.’ (MCCfn049/81)
- (12) *Inga Heideman: Csak annyi-t szeretné-k monda-ni,*
 Inga Heideman only so.many-ACC would.like-1SG say-INF
hogy vigyáz-z-anak az egészség-ük-re.
 that take.care-IMP-INDEF.3PL the health-POSS.3PL-SUB
 ‘Inga Heideman: All I want to say is to take care of your health.’ (formal; in a fake interview) (MCCfn111/112)

In the case of (11), the author of the text advertises the grid glasses through the mouth of a fictional professor, similarly to the occurrence in (12), where the directive force of a fictional interviewee’s utterance is mediated to the readers by publishing a fake interview. Professor Petz in example (11) and Inga Heideman in example (12) are fake authority figures who were employed to intensify the credibility of statement and avoid responsibility.

4.2 Utterances with a discourse function

Utterances that play a role in the creation and functioning of discourse were separated into the categories tagged as *szövegszervező* ‘discourse organizer’ or *interakció* ‘interactive’ functions. Even in contexts where these utterances fulfill a directive function, the expected action is often a mental activity related to the organization or understanding of the ongoing discourse or to the speaker-listener interaction, such as attention management and recalling and connecting information. In most cases, these forms occur as independent simple sentences or in a main clause, but they can also appear in subordinate clauses. In these latter cases the annotated function does not extend to the entire utterance. Some of these occurrences are discourse markers, but not all of them.

We assigned the “discourse organizer” tag to occurrences that have a textual function, such as creating discourse coherence or directing attention to different parts of the text, as in (13) and (14).

- (13) *Lás-s-uk, hogy történ-ik a befogadó sejt-ek megfertőzés-e.*
 see-IMP-DEF.1PL how happen-INDEF.3SG the recipient cell-PL infection-POSS.3SG
 ‘Let’s see how the infection of the host cells occurs.’ (MCCfn323/3)

- (14) *Természetesen a gyógyszer-ek, mégpedig pszicho-stimuláns drog-ok, mint a Ritalin, Concerta, Methylin, Equasym XL, hogy csak néhány márká-t említ-s-ek.*
 of.course the drug-PL namely psycho.stimulant drug-PL like the Ritalin Concerta Methylin Equasym XL that just some brand-ACC mention-IMP-INDEF.1SG
 ‘Of course, the drugs, namely psychostimulant drugs such as Ritalin, Concerta, Methylin, Equasym XL, to name just a few brands.’ (MCCfn678/67)

In example (13), the function of the utterance containing the imperative form is to organize the text via directing attention of the reader to the next topic. The utterance in (14) is a fragment from a fake news text where the interviewee puts forward a conspiracy theory about pharmaceutical companies, accusing them of deliberately diagnosing children as hyperactive in the USA in the 1990s. Example (14) is intended as an answer to the question ‘What was the solution?’. The addition *hogy csak néhány márkát említsek* ‘to name just a few brands’ with the imperative verb form embedded in a subordinate clause has a discourse-organizing function. Using the first person singular in this Hungarian formula, the speaker expresses a metapragmatic comment by naming their own discourse-organizing activity that facilitates understanding.

The *interakciós* ‘interactive’ label was assigned to those occurrences that are related to the activity of the speaker or the listener in connection with the creation and understanding of their discourse. These occurrences can refer to the process of meaning construction, the knowledge, feelings, and attitudes of the speaker or the listener. Consider (15) and (16):

- (15) *Gondol-j-anak bele!*
 think-IMP-INDEF.3PL PVB
 ‘Think about it!’ (formal) (MCCfn764/49)
- (16) *Hogy őszinte le-gy-ek, még soha nem hall-ott-am a Detonic-ról.*
 that honest be-IMP-INDEF.1SG yet never not hear-PAST-1SG the Detonic-DEL
 ‘To be honest, I’ve never heard of Detonic before.’ (MCCfn274/241)

In (15), the request to direct the reader’s attention appears as an independent directive utterance. In contrast, in (16), the imperative verb form is part of a subordinate clause expressing the speaker’s attitude, but the statement as a whole has no directive function.

4.3 Utterances that do not have directive or discourse functions

Utterances that contain an imperative verb form but do not have any directive or discourse function were tagged as *nodirectiva* ‘nondirective’ or *meta* ‘meta’. The “nondirective” label was assigned in cases where the use of the imperative form was present in utterances for grammatical reasons, with no directive function; the imperative form is obligatory in certain Hungarian linguistic structures (see Section 2). Verbs in imperative form of this type occur most often, although not exclusively, in subordinate clauses. In our corpus, the most common type were *that*-clauses expressing purpose, such as in (17).

- (17) *Könnyen le-het, hogy igaz-ak az-ok a hír-ek, hogy Kíná-ban Vuhan határ-á-ban rengeteg test-et éget-nek el an-nak cél-já-ból, hogy ne derül-hes-s-en ki, mekkora is a baj.*
 easy be-POT that true-PL that-PL the news-PL that China-INE Vuhan border-POSS.3SG-INE lot.of body-ACC burn-3PL PVB that-DAT goal-POSS.3SG-ELA
 that no turn.out-POT-IMP-INDEF.3SG PVB how.big too the trouble
 ‘The rumors that a lot of bodies are being burned on the outskirts of Wuhan in China to keep the extent of the problem under wraps can easily be true.’ (lit. ‘The rumors that a lot of bodies are being burned on the outskirts of Wuhan in China with the aim of not turning out how big the problem is, can easily be true.’) (MCCfn014/15)

The subordinate clause of purpose in (17) contains the imperative verb form *derülhessen ki* ‘to turn out’ that is required by the construction *annak a céljából, hogy* ‘with the aim of’.

In other “nondirective” utterances, the subordinate clause describes a desired state, the realization of something uncertain. Although the expression of a wish or want may have a directive force in certain contexts in Hungarian, in some other cases it does not have any imperative function, especially in the past tense, like in (18).

- (18) *A fiatal feltaláló azonban az-t szeret-t-e*
 the young inventor however that-ACC like-PAST-3SG
volna, hogy ez a forradalmi felfedezés
 would that this the revolutionary discovery
ne csak a gazdag-ok számára le-gy-en
 not only the rich-PL for be-IMP-INDEF.3SG
elérhető, hanem bárki használ-has-s-a.
 available but anyone USE-POT-IMP-DEF.3SG

‘However, the young inventor wanted this revolutionary discovery not only to be available to rich people, but for anyone to use.’ (MCCfn049/20)

The use of the imperative form is also obligatory in some Hungarian linguistic structures expressing possibility/impossibility/necessity or prohibition/permission.

- (19) *Emellett nincs mód-j-uk ar-ra sem, hogy fel-mond-j-ák*
 besides not.be.3SG way-POSS-3PL that-SUB neither that PVB-terminate-IMP-DEF.3PL
közalkalmazotti munkaszerződés-eik-et.
 civil.servant contract-POSS.3PL.PL-ACC
 ‘In addition, they have no way to terminate their employment contracts as civil servants.’ (MCCpmo147/6)

In addition to these most common types, the imperative form can appear in many other subordinate structures in nondirective utterances in the Hungarian language. Only a small proportion of imperative forms occurred in simple sentences. If these did not have a directive function, they were also labeled as “nondirective”, like in (20):

- (20) *Ve-gy-ek be kalcium-ot allergiás reakció esetén?*
 take-IMP-INDEF.1SG PVB calcium-ACC allergic reaction in.case.of
 ‘Should I take calcium in case of an allergic reaction?’ (MCCpir015/17)

We assigned the “meta” tag when an imperative form occurs in an utterance that describes or invokes a directive, but which is not directed towards the reader as a potential addressee. The most common case of this is when the directive is not relevant for current readers either because it happened in the past and was performed to a different addressee, as in (21), or because it is addressed to someone else, like in (22), where it is directed towards the interviewee in an interview.

- (21) *Haza-küld-t-ék az-zal, hogy ha rosszabbul van, jöj-j-ön vissza.*
 to.home-send-PAST-3PL that-INS that if worse be.3SG come-IMP-INDEF.3SG PVB
 ‘She was sent home with the instruction to come back if she got worse.’ (MCCfn348/8)
- (22) *Mond-j-a el, mi-t kell ten-ni-e egy személy-nek a Keramin krém*
 tell-IMP-DEF.3SG PVB what-ACC must do-INF-3SG a person-DAT the Keramin cream
megszerzés-é-hez.
 acquisition-POSS-ALL
 ‘Tell us what a person needs to do to get Keramin Cream.’ (formal) (MCCfn763/133)

Example (21) is a part of a longer story about a sick girl, where the narrator recalls what the doctor told the girl. The directive was addressed to the girl in the past; it does not apply to current readers. The directive in (22) is a request addressed to the interviewee by the interviewer, and once again the directive force is not relevant to potential readers.

In other cases, the author of the text distances themselves from the quoted directive and does not convey its imperative force to the readers, even if this often only becomes clear by considering a wider context, as seen in (23):

- (23) *Senki-t sem szabad orvosi kísérlet-re kényszerít-eni informált beleegyezés*
 nobody-ACC neither allowed medical experiment-SUB force-INF informed consent
nélkül. Sok média, politikai és nem orvosi személy
 without many media political and not medical person
az-t mond-ja az ember-ek-nek, hogy men-j-enek el az
 that-ACC tell-DEF.3SG the man-PL-DAT that go-IMP-INDEF.3PL PVB the
oltás-ra, az biztonságos, és nem nyújt-anak
 vaccination-SUB that safe and not provide-INDEF.3PL
információ-t a gén-terápia káros hatás-ai-ról vagy
 information-ACC the gene.therapy harmful effect-POSS.3SG.PL-DEL or
veszély-ei-ről.
 danger-POSS.3SG.PL-DEL
 ‘No one should be forced to undergo a medical experiment without informed consent. Many media, politicians, and nonmedical experts tell people to go for vaccination, it is safe, and they do not provide information about the adverse effects or dangers of gene therapy.’ (MCCfn319/33–34)

The wider context of (23) reveals that the narrator adopts a dismissive attitude towards the views of the media, politicians, and nonmedical persons they refers to.

5 Results of the statistical analysis and discussion

Having presented our classification of utterances, we can now turn to the statistical analysis of our results concerning the functions of utterances with imperative forms in fake and real news. As it was mentioned above, during the manual pragmatic annotation of the MedCollect corpus, annotators assigned labels to words that had been identified as imperative verb forms based on the preliminary automatic morphological analysis. In the subcorpus containing a total of 366,832 tokens, there were 2,542 imperative forms identified. The number of occurrences of imperative forms in various functions within fake news and real news texts are provided in Table 1.

It can be observed that there are significantly more imperative forms, particularly “own voice” directives, in the fake news texts, despite the total number of tokens in the analyzed fake news being lower than in the real news texts. The ratio of imperative forms per 10,000 tokens allows for a more precise comparison, as can be seen in Table 2. In fake news texts, imperative forms occur more than twice as frequently as in real news texts. Furthermore, the frequency of “own voice” directives is even higher, with more than ten times as many. Indeed, the frequency of “own voice” directives in fake news articles alone is nearly equivalent to the total frequency of

Table 1: Imperative forms and their functions.

	Fake news	Real news	Total
Tokens	162,391	204,441	366,832
Imperative forms	1,645	897	2,542
Own voice	804	88	892
Mediated	81	103	184
Discourse organizer	44	25	69
Interactive	36	10	46
Nondirective	631	577	1,208
Meta	49	94	143

Table 2: The relative frequency of imperative forms and their functions per 10,000 tokens.

	Fake news	Real news
Imperative forms	101.3	43.9
Own voice	49.5	4.3
Mediated	5.0	5.0
Discourse organizer	2.7	1.2
Interactive	2.2	0.5
Nondirective	38.9	28.2
Meta	3.0	4.6

Table 3: The ratio of different functions of imperative forms relative to all imperative forms in fake and real news.

	Fake news	Real news
Own voice	0.489	0.098
Mediated	0.049	0.115
Discourse organizer	0.027	0.028
Interactive	0.022	0.011
Nondirective	0.384	0.643
Meta	0.030	0.105

imperative forms found in real news texts. These data support our hypothesis that the use of imperative forms is more characteristic in fake news articles, even without distinguishing between functions.

Table 3 shows the relative frequency of different functions of imperative forms relative to all imperative forms. Here, it can also be observed that imperative forms in fake news express “own voice” directives much more frequently than in real news texts (approximately five times more often). In contrast, real news texts demonstrate a higher frequency of “mediated” and “meta” directives as well as “nondirective” functions, categories of utterances that do not perform directives. As for the “discourse organizer” and “interactive” functions, they occur in the same proportion in both types of texts.

These data support the hypothesis that fake news texts more frequently employ strategies directly putting pressure on the reader and portraying the text’s author as the source of experience. In real news, however, it is more common to mediate the directives of others, such as experts.

6 Conclusions

The current corpus-based, comparative statistical analysis of the use of imperative forms in Hungarian health-related fake and real news has shown that the theoretically based, fine-grained empirical identification of the linguistic differences between fake and real news can contribute to the detecting of fake news. The methodology employed in the research – that is, the combination of automatic (morphological) analysis with manual pragmatic annotation – as well as the combination of qualitative and quantitative analyses has proved to be a fruitful tool for detecting disinformation. This methodology could be applied both to similar studies in other languages and to other linguistic features which characterize fake news, and not only in the field of healthcare. Consequently, this type of study could be a very useful tool in the fight against the dissemination of false information.

Research funding: Our research was supported by the Subprogramme for Linguistic Identification of Fake News and Pseudo-scientific Views, part of the Science for the Hungarian Language National Programme of the Hungarian Academy of Sciences (2022–2026, <http://enyik.szte.hu/>, <https://www.alhirdetektor.hu/>).

References

- Aijmer, Karin. 2018. Corpus pragmatics: From form to function. In Andreas H. Jucker, Klaus P. Schneider & Wolfgang Bublitz (eds.), *Methods in pragmatics*, 555–586. Berlin: De Gruyter Mouton.
- Bencédy, József, Pál Fábrián, Endre Rácz & Mártonné Velcsov. 1976. *A mai magyar nyelv* [The contemporary Hungarian language]. Budapest: Tankönyvkiadó.
- Dömötör, Éva & Diána Varga. 2014. Főmondati kötőmód diskurzuspartikulákkal [Main clause subjunctive with discourse particles]. In Zsuzsanna Gécség (ed.), *LingDok 13: Nyelvészdedoktoranduszok dolgozatai* [LingDok 13: Papers of PhD students in linguistics], 27–45. Szeged: Szegedi Tudományegyetem.
- Eckart de Castilho, Richard, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank & Biemann Chris. 2016. A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the workshop on language technology resources and tools for digital humanities (LT4DH)*, 76–84. Osaka: COLING 2016 Organizing Committee.
- É. Kiss, Katalin. 1998. Mondattan [Syntax]. In Katalin É. Kiss, Ferenc Kiefer & Péter Siptár (eds.), *Új magyar nyelvtan* [New Hungarian grammar], 141–142. Budapest: Osiris.
- Grieve, Jack & Helena Woodfield. 2023. *The language of fake news*. Cambridge: Cambridge University Press.
- Hegedűs, Rita. 2004. *Magyar nyelvtan: Formák, funkciók, összefüggések* [Hungarian grammar: Forms, functions, connections]. Budapest: Tinta Könyvkiadó.
- H. Varga, Márta. 2020. Kötőmód a magyarban és némely indoeurópai nyelvben [Subjunctive in Hungarian and some Indo-European languages]. *Publicationes Universitatis Miskolcensis, Sectio Philosophica Tomus 24*(1). 214–225.
- Keszler, Borbála (ed.). 2000. *Magyar grammatika* [Hungarian grammar]. Budapest: Nemzeti Tankönyvkiadó.
- Kiefer, Ferenc (ed.). 2018. *Strukturális magyar nyelvtan 3: Morfológia* [Structural Hungarian grammar 3: Morphology]. Budapest: Akadémiai Kiadó.
- Kugler, Nóra. 2000. Az igeragozás [Verbal conjugation]. In Borbála Keszler (ed.), *Magyar grammatika* [Hungarian grammar], 104–126. Budapest: Nemzeti Tankönyvkiadó.
- Péteri, Attila. 2012. The Hungarian imperative particle *hadd*: A contrastive look at the development of analogous markers of sentence mood in some European languages. *Acta Linguistica Hungarica* 59(4). 439–463.
- Prileszky, Csilla. 1974. A felszólító módú mellékmondat néhány kérdéséről [On some issues of the imperative clause]. *Nyelvtudományi Értesítő* 83. 473–475.
- Scott, Kate. 2021. You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics* 175. 53–66.
- Searle, John R. 1979. A classification of illocutionary acts. *Language in Society* 6. 1–23.
- Simonyi, Zsigmond. 1895. *Tüzetes magyar nyelvtan történeti alapon* [Accurate Hungarian grammar on a historical basis]. Budapest: Akadémiai Kiadó.
- Sousa-Silva, Rui. 2022. Fighting the fake: A forensic linguistic analysis to fake news detection. *International Journal for the Semiotics of Law* 35. 2409–2433.
- Szarka, Péter. 2000. Van-e kötőmód a magyarban? [Are there subjunctives in Hungarian?]. *Hungarológia* 1–2. 249–259.
- Szécsényi, Tibor, Katalin Nagy C. & Németh T. Enikő. 2024. Felszólításannotálás a MedCollect egészségügyi álhírkorpuszban [Annotation of the imperative forms in MedCollect healthcare fake news corpus]. In Gábor Berend, Gábor Gosztolya & Veronika Vincze (eds.), *XX. Magyar Számítógépes Nyelvészeti Konferencia* [20th Conference of Hungarian Computational Linguistics], 159–170. Szeged: Szegedi Tudományegyetem.
- Tompa, József (ed.). 1962. *A mai magyar nyelv rendszere I–II* [The system of today's Hungarian language, vols. 1–2]. Budapest: Akadémiai Kiadó.
- Tóth, Enikő. 2003. A magyar kötőmódról [About the Hungarian subjunctive]. In István Hoffmann & Tamás Kis (eds.), *Magyar Nyelvjáráskok XLI*, 617–626. Debrecen: Debreceni Egyetem Magyar Nyelvtudományi Tanszék.
- Tóth, Enikő. 2014. The imperative and the subjunctive proper: Two distinct grammatical moods in Hungarian. *Argumentum* 10. 631–644.
- United Nations. 2023. Information integrity on digital platforms (our common agenda policy brief 8). United Nations. <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf> (accessed 16 November 2023).
- Vincent, Benet, Kate Power, Peter Crosthwaite & Sheena Gardner. 2023. Directives in COVID-19 government guidance: An international comparison. *Applied Corpus Linguistics* 3(4). <https://doi.org/10.1016/j.acorp.2023.100063>.
- Zhou, Zhixuan, Huankang Guan, Meghana Moorthy Bhat & Justin Hsu. 2019. Fake news detection via NLP is vulnerable to adversarial attacks. *Proceedings of the 11th international conference on agents and artificial intelligence, vol. 2, ICAART*, 794–800. Prague: Springer.