ENCODING DIRECT ADDRESS BEYOND THE USUAL SUSPECTS. A CROSS-LINGUISTIC STUDY

Georgi Georgiev

Sofia University "St. Kliment Ohridski"

ORCID ID: 0009-0001-7632-4445

email: stilijanog@uni-sofia.bg

Abstract: In language, the same content may be encoded in different ways. For instance, participant relations can be encoded using case, word order and adpositions. This article examines a sample of 89 languages that do not seem to display morphologically marked case (m-case), including a vocative, on full nouns (or NPs). The principal hypothesis put forward is that direct address can be encoded in ways that go beyond the vocative phrase, the vocative and intonation contours. Having presented a brief overview of key notions, the article surveys the data from the sample and discusses them. Relevant conclusions are drawn.

Keywords: (direct) address, addressee, encoding, vocative, vocative phrase, particles

Резюме: Известно е, че в езика едно и също съдържание може да намери различен израз. Например актантните отношения могат да бъдат кодирани чрез падежи, словоред, предлози и следлози. В настоящата статия се разглежда извадка от 89 езика, в които при съществителните имена (или именните фрази) не се откриват морфологично маркирани падежи, включително вокатив. Формулира се основната хипотеза, че прякото обръщение може да намери израз, различен от звателната фраза, вокатива и интонационните контури. След кратък преглед на основните понятия се пристъпва към обзор на данните от извадката и коментар по тях. Представят се относими заключения.

Ключови думи: (пряко) обръщение, адресат, кодиране, вокатив, звателна фраза, частици

A well-known truism in linguistics states that the same content may be encoded differently in the same language and, indeed, in distinct languages. A case in point are participant relations. These may be encoded using morphologically marked case (henceforth, *m-case*), word order and adpositions among others (Blake 2004: 12-17; Butt 2006: 4 *et seq.*). Taking this observation as a point of departure, in this article I report on data coming from a sample of 89 languages with no m-case, including no vocative, on full nouns (or NPs) and look at how direct address is encoded in these beyond the vocative phrase.

This article is structured as follows: section 1 traces prominent notions of the vocative and direct address; section 2 spells out the main hypothesis and presents the limitations of this study; section 3 surveys the data collected and discusses these in brief; section 4 offers concluding remarks. An appendix lists the languages in the sample.

1. Direct address and its encoding

Address is a broad notion. It may be direct or indirect. Direct address is used to attract and maintain the attention of the addressee by identifying them directly (*John, would His Majesty like anything else?*). For its part, indirect address does not identify the addressee directly and instead relies on generally sociolinguistic or pragmatic conventions for this. In *Would His Majesty like anything else?* the addressee could be a hypothetical king. This would thus be an example of indirect address since the addressee can only infer that he is being addressed based on his knowledge of the convention at issue. Alternatively, the addressee could be a servant in the same room. The servant's status as an addressee, however, will have to be inferred from context since there is no way to identify them as such based solely on what has been said. That is because in this particular situation *His Majesty* refers to an individual present in the room that is not the addressee. Where the addressee's identity has be inferred solely from co(n)text, there is no address, be it direct or indirect.

The notion of address should be distinguished from that of exclamation, after Hill (2014: 6–8). This is reflected in Daniel and Spencer (2009: 626), who talk of 'calling out [= exclamation] and attracting or maintaining the addressee's attention.' While address always requires an addressee, exclamation does not. Thus, when addressing God directly in Bulgarian, one may utter *Боже, дай ми сили!* ('God, give me strength!'), but one can just as easily utter *Боже!* ('God!') to express frustration without actually (thinking that they are) addressing anyone. For the purposes of this article, I shall use a narrow understanding of direct address. This will be considered to involve i. directly attracting and/or maintaining the attention of an addressee (real or imagined, present or absent, etc.) and ii. doing this by identifying them through denotation, reference and/or classification (in terms of their number or other variables, for instance).[1]

The literature identifies multiple means of encoding direct address. One is the vocative. Work investigating the vocative from a cross-linguistic perspective is not too ample. In fact, there are few quantitative studies on it and qualitative ones are not too numerous either (Daniel and Spencer 2009; Janson 2013; Hill 2014). Blake (2004: 8) states that the vocative 'is used as a form of address' – an assertion that needs to be qualified. In the case at hand, Blake means that the vocative can encode direct address. The vocative can also encode exclamation. In that case, it is used in formulaic expressions. Aligning with Blake (2004), Donati (2013: 270) considers that the vocative serves to '[place] an addressee in a given speech context' and highlights the fact that it identifies the addressee directly. Taken together, the stances above suggest that the vocative encodes direct address consistently and may further encode exclamation but not as consistently.

In view of the above, it is necessary to distinguish between the vocative phrase and the vocative. The understanding of the term *vocative phrase* adopted here follows Hill (2014: 4–8). A vocative phrase is organised around a nominal (including nominalised adjectives and pronouns), which may or may not be accompanied by other elements (*добри човече* ('good person') in direct address, for instance). Direct address is always encoded by the vocative phrase and may be encoded by the vocative in parallel, but only when the vocative appears in

---

[1] Sonnenhauser and Noel Aziz Hanna (2013: 13–15) provide a concise summary of the debate around the content encoded by the vocative phrase, the vocative and a set of related phenomena.

the vocative phrase. Whether the vocative encodes direct address or exclamation is always conditioned by its immediate syntactic environment, i.e. whether it is in the vocative phrase[2], but the vocative still encodes this content in parallel with the phrase in which it appears. Thus, where the vocative is found in the vocative phrase, there is cumulation of distinct means of encoding direct address.

Positing two separate means of encoding – one syntactic and the other morphological – for the same content and stating that these can operate in parallel is a claim that has to be demonstrated to be true rather than assumed. To do so, let us consider a set of examples. In Bulgarian, *Това си ти, Иван.* ('This is you, Ivan.') is ambiguous when put down in writing. It could be interpreted as a case of apposition where the personal pronoun and *Иван* both identify the same individual as the addressee but where the speaker does not seek to attract or maintain the addressee's attention. An alternative interpretation would be to treat *Иван* as a nominal appearing in the vocative phrase, i.e. in direct address. In that case, the pronoun and *Иван* both identify the addressee, but *Иван* being in the vocative phrase also encodes the fact that the speaker seeks to attract or maintain the addressee's attention. Similarly, two alternatives are available when we have *Иван, това си ти.* ('Ivan you are.' or 'This is you, Ivan,' the latter encoding direct address). In this example, *Иван* could be interpreted as being fronted and thus thematised. This corresponds to the first translation above. The direct address interpretation, which corresponds to the second translation, was already commented upon, so we need not repeat that.

The cases of ambiguity above stem from the fact that *Иван* is not in the vocative in either example. In contrast, we can also have *Това си ти, Иване.* ('This is you, Ivan.' – direct address) and *Иване, това си ти.* ('Ivan, this is you.' – direct address), where we do have the vocative. Since these are not formulaic expressions, an exclamation interpretation is ruled out, which means that the vocative can only encode direct address in both examples. This demonstrates that the encoding of direct address is both at the level of syntax and at that of morphology. A similar argument can be made about intonation. All languages have a vocative phrase, but some also need a vocative, and many most likely employ intonation contours that may be more or less specific with the vocative phrase. It is worth noting that a number of the grammars consulted state that the intonation contour typical of a vocative phrase is not always in place when the vocative phrase is not in its canonical position for a given language. In such cases, a different, often less distinctive contour may replace the typical one. It is in these cases that the vocative is particularly informative. If we circle back to our original four examples and consider only their direct address interpretations, we can observe that direct address is encoded as follows:

Table 1. Encoding of direct address in a set of sentences in Bulgarian

IN ORAL DISCOURSE		IN NON-ORAL DISCOURSE	
Това си ти, Иван.	syntax (word order) intonation	*Това си ти, Иван.*	syntax (word order)
Иван, това си ти.	syntax (word order) intonation	*Иван, това си ти.*	syntax (word order)

---

[2] Using the movability test, Hill demonstrates that the distinction between direct address and exclamation is also encoded in syntax, i.e. that direct address and exclamation are encoded in two distinct syntactic constituents (direct address – in the vocative phrase, and exclamation – in another constituent) that may, in fact, co-occur.

Това си ти, Иване.	syntax (word order) intonation vocative	*Това си ти, Иване.*	syntax (word order) vocative
Иване, това си ти.	syntax (word order) intonation vocative	*Иване, това си ти.*	syntax (word order) vocative

The vocative phrase is used to encode direct address in all cases. However, specific intonation contours are used in oral discourse and the vocative may also be used. In non-oral discourse, the examples in the grey boxes are ambiguous. For them to be interpreted as featuring a vocative phrase, co(n)text needs to be invoked. This ambiguity is resolved when there is a vocative, however.

Above, I singled out the vocative phrase as the most prominent means of encoding direct address. Two additional means were identified – the vocative and dedicated intonation contours. Both of these are associated with the vocative phrase. A further two means have to be mentioned. One is particles that accompany nominals in the vocative phrase. These may be optional or obligatory and pre- or post-posed, as well as pre- and post-posed at the same time. They can also display restrictions as to the nominals that they accompany. Such particles are only considered means of encoding direct address if they identify the addressee in addition to attracting or maintaining their attention. While some of the 89 languages in the sample discussed below appear to have particles accompanying nominals in the vocative phrase, these particles do not appear to identify the addressee in any of these languages. Similar particles that do in fact identify the addressee will receive discussion in section 3.2. Dedicated terms of address[3] are another means of encoding direct address. At least 9 of the 89 languages in the sample seem to feature such terms. Thus, it appears that direct address tends to be encoded by various means (the vocative phrase, the vocative, intonation contours, adnominal address particles and dedicated terms of address) and that these are often used cumulatively. In spoken Bulgarian, for instance, one can conceivably use up to four of these at the same time. That being said, the vocative phrase is the most prominent means of encoding direct address.

An explanation for the observation above is that direct address is likely highly salient in ontological terms. It thus bears asking whether there might not be means of encoding it beyond the ones mentioned earlier.

2. Hypothesis and limitations

Taking the potentially high ontological salience of direct address as a starting point, I propose that direct address is not solely encoded by the vocative phrase and the means of encoding associated with it. I aim to demonstrate that its means of encoding are not homogenous and can extend beyond the vocative phrase. I suggest that some means of encoding allow the

---

[3] Dedicated terms of address are lexemes that are only used in the vocative phrase and correspond to lexemes with the same denotation used outside the vocative phrase. Dedicated terms of address lexically encode that the attention of the addressee they denote is being attracted or maintained. Where such lexemes appear outside the vocative phrase, for instance in exclamations, they cannot be considered terms of address anymore. That is because their use is conventionalised, does not involve attracting or maintaining the addressee's attention and is accompanied by a loss of their capacity to identify an addressee lexically.

addressee whose attention is attracted or maintained to be identified with greater ease than others do.

To test if this is so, I look at a convenience sample of 89 typologically distinct, areally dispersed languages. These languages largely lack close genetic ties and do not appear to have m-case, including the vocative.

The survey reported on below is limited in terms of scope. Its discussion does not focus on the make-up of the vocative phrase. Instead, it concerns phenomena observed outside it and potentially in parallel to it. The survey is also limited in terms of depth. Many of the descriptive grammars consulted offer little information on the encoding of direct address. However, as observed, direct address is likely ontologically salient. Hence, it is unlikely that all the languages that seem to lack means of encoding it beyond those cited in the previous section do, in fact, lack these. It is more probable that such means did not make it into the relevant grammars. This constitutes an important limitation since it precludes a fully-fledged quantitative assessment. Thus, I shall consider the data collected only in relative terms because the data still allow for rough generalisations as to the comparative prevalence of phenomena across the languages in the sample. In this way, I aim to ensure the reliability of my conclusions and avoid potentially misrepresenting the languages whose grammars do not describe the encoding of direct address in great detail.

3. Data and discussion

The sample used for the survey reported on below covers 89 languages that form part of a larger sample presented in Iggesen (2013). All languages come from the group lacking m-case. Since Iggesen does not consider the vocative an m-case value, it was necessary to collect data independently for the purposes of our sample. The 89 languages covered in our sample have been identified as lacking the vocative.[4]

Below, I present an overview of the data collected (subsection 3.1) and I briefly discuss these (subsection 3.2).

3.1. Survey data

Of the languages surveyed, only three can be said to display convincing evidence suggesting that direct address is encoded beyond the vocative phrase, i.e. that the attention of the addressee is directly attracted or maintained and that this is done by identifying the addressee. In two of these languages, this is achieved through independent direct address particles. The grammar of each language states that the particles do not have nominal characteristics (they cannot refer or denote, for instance) and only serve to classify addressees, i.e. to limit their potential range. This is corroborated by the data presented in said grammars. In a third language, while the particles are described as being independent, evidence seems to point in the direction that the phenomenon in question is an instance of head-marking. Below, I consider each language in turn.

3.1.1. Lavukaleve

---

[4] Data points have not been cited in this article due to word limit considerations and can be made available upon request.

Lavukaleve is an endangered language spoken on the Solomon Islands. It employs a set of utterance-final particles to encode the number of addressees (*na* – one addressee, *nail* – two addressees, *nai* – more than two addressees). These are typical of, but not confined to, commands and exhortations and they are used in very informal interaction. The particles are preceded by an intonation break and 'always receive very strong stress' (Terrill 2003: 62–63). As mentioned earlier, they do not display any nominal characteristics and, judging by the examples adduced in the source consulted, they serve to focus the addressee's attention[5] on the part of the utterance that precedes them:

Lavukaleve; adapted from (Terrill 2003: 63)

Hamus ho'bea, na!

hamus   ho'bea   na

evening(N)[SG]   good[N.SG]   SINGLE.ADDRESSEE

Good night, you! (the *you* is a functional equivalent – *na* is not a pronoun)

The addressees are identified less readily using this particle than they are when a nominal is used in the vocative phrase. That is because they are only classified based on their number.

3.1.2. Wari'

Wari' is an endangered language spoken chiefly in a region of the Amazon along the border between Brazil and Bolivia. The vocative phrase in Wari' generally appears in sentence-final position (Everett and Kern 1997: 244–245). There are also two emphatic particles roughly translated as 'hey, you.' Data suggest that these do not display any nominal characteristics. These particles serve to attract the addressee's attention and are generally used in isolation (Everett and Kern 1997: 195–196). One (*wira*) is used when men address men and the other (*cama*) – when women address women. In a situation where there are multiple potential addressees, some of them male and others – female, the use of these particles enables the partial identification of addressees and, thus, the direct attraction or maintenance of their attention. Again, the addressees are identified less readily than when a nominal is used in the vocative phrase. That is because they are only identified on the basis of whether they are male or female.

3.1.3. Lakhota

Lakhota is an endangered language spoken chiefly in North and South Dakota in the USA. To express commands, permissions and mild requests, Lakhota employs particles that are post-posed with respect to the verb. These particles encode whether we are dealing with a command, a permission or a mild request, whether this comes from a man or a woman and the number of addressees intended to take action. Again, the particles do not display any nominal characteristics and all of them are accented except for one, which is often omitted (Boas and Deloria 1941: 111–112).

The way in which the particles in Lakhota behave (they differ depending on whether we have a command, a permission or a mild request) suggests that they are closely related to the verb phrase. They are strictly confined to specific speech acts and focus the addressee's attention

---

[5] Lavukaleve has an utterance-initial attention-getter, which does not identify the addressee (Terrill 2003: 62).

by identifying them. However, the attention of the addressee is directed towards the verbal action, not towards the entire utterance, which is a departure from what we saw in Lavukaleve and Wari'. As Bozhil Hristov (p.c.) suggests, taken together, the considerations listed in this paragraph point to the phenomenon in Lakhota likely being an instance of head-marking rather than a fully-fledged set of independent direct address particles. This is in line with the data pointing to Lakhota being a head-marking language throughout Boas and Deloria (1941).

3.2. Discussion

Let us consider the quantitative part of our data first. The 89 languages in the sample were selected because they do not seem to have the vocative (and m-case in general). Hence, the vocative cannot be included in our discussion below. The same goes for intonation contours since the majority of the sources consulted lack sufficient information on these. This leaves us with the data below:

Table 2. Select means of encoding direct address in a sample of 89 languages lacking m-case, including the vocative

MEANS OF ENCODING	NUMBER OF LANGUAGES
vocative phrase	89/89
dedicated terms of address	at least 9/89
independent direct address particles	at least 2/89

Expectedly, all languages in the sample have a vocative phrase. Other means of encoding are far less commonly attested. In addition, independent direct address particles are limited in their distribution in distinct ways. The particles in Wari' are generally used in isolation, while those in Lavukaleve are utterance-final. That said, the scarcity of data available does not allow for generalisations concerning the contexts of independent direct address particles.

Turning to qualitative considerations, the first two means of encoding in the table require a nominal. Thus, direct address is encoded thanks to the interaction of denotation, reference and/or classification (for instance, if a single addressee is meant, the singular will likely be used). More often than not, at least two out of the three – denotation, reference and classification – are used. This is not so with independent direct address particles. They serve to attract or maintain the attention of addressees by identifying them partially and solely through classification. In the languages examined, they limit the range of potential addressees by classifying them in terms of their number (Lavukaleve)[6] or whether they are male or female (Wari').

There are languages in which address particles may be independent or accompany nominals in the vocative phrase. This can only be ascertained by looking beyond the data collected and discussed above. In Bulgarian, which is not part of our sample because it has the vocative, the particle *ма* may appear utterance-finally to identify a single female addressee in informal interaction: *Той ще иде до магазина, ма.* ('He will go to the shop.' – addressed to a single female addressee). In addition, it may accompany a nominal in the vocative phrase and appear pre- or post-posed, as well as pre- and post-posed at the same time, and classify the addressee in the same way: *Ма Женьо/Женьо ма/Ма Женьо ма, стига си се оплаквала!* ('Hey,

---

[6] In Lakhota, addressees are classified based on their number. Due to the uncertain status of the phenomenon in Lakhota, I have opted to leave this out of the generalisation above.

Zhenya, stop complaining already!').[7] What is more, preliminary data from Bulgarian indicate that, when *ма* is used with a noun, this noun has to be in the vocative (in cases where there is a vocative such as the one above).

The observations above suggest that attracting and maintaining the addressee's attention through direct identification constitute a common thread linking all means of encoding direct address. How identification is achieved tends to differ. The vocative phrase and the means of encoding associated with it require the presence of a nominal and rely on different combinations of denotation, reference and classification to identify the addressee to a greater or lesser degree. This is not unexpected since it is nominals that contribute both the denotational and the referential meaning. Classification in such cases is generally observed to operate directly on the nominal (a nominal may be marked for number, etc.). Direct address particles may be independent or accompany a nominal. In either case, they always identify the addressee partially through classification.

4. Concluding remarks

The discussion above suggests several avenues for furthering work on the means of encoding direct address. For instance, it would be worthwhile to explore the interaction of these means in the languages in the sample insofar as such interaction is attested. It would also be useful to look beyond the sample. Data on languages that have the vocative need to be collected in view of the observations on Bulgarian in subsection 3.2. In addition, it could be that languages that feature specific means of encoding direct address have a greater likelihood of featuring others. Hence, this is also worth investigating.

References:

Blake 2004: Blake, B. J. Case. Cambridge: Cambridge University Press, 2004.
Boas and Deloria 1941: Boas, F. and Deloria, E. Dakota Grammar. Washington: Government Printing Office, 1941.
Butt 2006: Butt, M. Theories of Case. Cambridge: Cambridge University Press, 2006.
Čašule 2021: Čašule, I. Macedonian Appellative Particles in the Balkan Context. – Balkanistica, 34/2021, 171–197.
Croitor and Hill 2013: Croitor, B. and Hill, V. Chapter 15. Vocatives. – In: A Reference Grammar of Romanian. Volume 1: The Noun Phrase. Amsterdam / Philadelphia: John Benjamins, 2013, 801–826.
Daniel and Spencer 2009: Daniel, M. and Spencer, A. The Vocative – An Outlier Case. – In: The Oxford Handbook of Case. Oxford: Oxford University Press, 2009, 626–634.

---

[7] It has been argued that the so-called *appellative particles* (which Hill (2014) terms *vocative particles*) – a cover term that subsumes direct address particles and attention-getters – are a feature of the Balkan language area (Hill 2014). Tisheva and Rå Hauge (2002) explore a broader set of *pragmatic particles*. These are said to encode 'the speaker's attitude towards various elements of the communicative situation' (Tisheva and Rå Hauge 2002: 11; translation mine). Appellative particles are included in this set. Owing to the broad scope of the terms *appellative particles* and *pragmatic particles* and the heterogeneity of the phenomena subsumed under these, direct address particles cannot automatically be considered a feature of the Balkan language area.
The considerations listed notwithstanding, above I observed that Bulgarian has direct address particles. In addition, Romanian has been reported to have such particles that behave in a similar way to those in Bulgarian (Croitor and Hill 2013: *passim*). The same goes for Macedonian (Čašule 2021: *passim*). Thus, preliminary evidence suggests that direct address particles could indeed be a feature of the Balkan language area in their own right – an issue that merits further investigation.

Donati 2013: Donati, M. The vocative case between system and asymmetry. – Vocative!: Addressing Between System and Performance. Berlin, Boston: De Gruyter Mouton. 2013, 269–282.

Everett and Kern 1997: Everett, D. L. and Kern, B. Wari'. London: Routledge, 1997.

Hill 2014: Hill, V. Vocatives: How Syntax Meets with Pragmatics (with the participation of Melita Stavrou). Leiden: Brill, 2014.

Iggesen 2013: Iggesen, O. A. Number of Cases. – In: WALS Online, 2013. <https://wals.info/chapter/49> [viewed 5.6.2024]

Janson 2013: Janson, T. (2013). Vocative and the grammar of calls. – In: Vocative!: Addressing between System and Performance. Berlin, Boston: De Gruyter Mouton, 2013, 219–234.

Sonnenhauser and Noel Aziz Hanna 2013: Sonnenhauser, B. and Noel Aziz Hanna, P. – In: Vocative!: Addressing between System and Performance. Berlin, Boston: De Gruyter Mouton, 2013, 1–23.

Terrill 2003: Terrill, A. A Grammar of Lavukaleve. Berlin: Mouton de Gruyter, 2003.

Tisheva and Rå Hauge 2002: Tisheva, Y. and Rå Hauge, K. Zaemane na pragmatichni chastitsi. – In: Problemi na sotsiolingvistikata, t. 7. Sofia: Delphi, 2002, 10–17. [*Тишева и Ро Хауге 2002*: Тишева, Й. и Ро Хауге, Х. Заемане на прагматични частици. – В: Проблеми на социолингвистиката, т. 7. София: Делфи, 2002, 10–17.]

Appendix: Sample of 89 languages lacking m-case, including the vocative

Abipón	Fula (Cameroonian)	Koromfe	Paiwan
Acoma	Grebo	Kosraean	Palauan
Amele	Guaraní	Koyraboro Senni	Rapanui
Arabic (Egyptian)	Haida	Kutenai	Sango
Arapesh (Mountain)	Hausa	Lakhota	Sentani
Asmat	Hebrew (Modern)	Lango	Slave
Bagirmi	Hixkaryana	Lavukaleve	Spanish
Bambara	Hmong Njua	Luvale	Swahili
Batak (Karo)	Huave (San Mateo del Mar)	Madurese	Taba
Canela-Krahô	Iaai	Malagasy	Tagalog
Carib	Igbo	Mandarin	Thai
Catalan	Indonesian	Maori	Tinrin
Chamorro	Italian	Marind	Tiwi
Chinantec (Lealao)	Jakaltek	Maung	Tsimshian (Coast)
Chumash (Barbareño)	Ju|'hoan	Mixtec (Chalcatongo)	Tuvaluan
Diola-Fogny	Kayah Li (Eastern)	Mokilese	Una
Dong	Khasi	Navajo	Urubú-Kaapor
Drehu	Khmer	Ndyuka	Vietnamese
Dutch	Khmu'	Nkore-Kiga	Wari'
Ewe	Kilivila	Oneida	Wichí
Fijian	Kiribati	Otomí (Sierra)	Yoruba
French	Kobon	Paamese	Zuni
Frisian (Western)			