

5 ICLEI 2016-57 Renata Šimůnková

A Comparative Study of English and Czech Verb Forms Based on a Parallel Corpus

Renata Šimůnková
Department of the English Language, Technical University of Liberec,
Studentská 1402/2, Liberec 1
renata.simunkova@tul.cz

ABSTRACT

The area of modality represents an extensive and complex issue and since the fundamental means of its expression differ from language to language, it often causes problems in translations between languages. On the other hand modals verbs are frequently used means of everyday language and a deeper insight into the differences between the systems is always very valuable. Parallel corpora represent efficient means for such studies as they allow a large number of tokens to be processed, which would be extremely demanding if performed by means of manual excerption, and they thus grant more objective results. The subject-matter of the paper is a study of the means used to express negative necessity (lack of permission, logical necessity not to) in English and their counterparts in Czech. In particular, by means of co-text analysis, different semantic values of these means of expression are delimited. The study further focuses on how the choice of particular means is influenced by the style of a text, the two styles investigated being fiction and journalese. The generally assumed distinction between MUST NOT and CAN NOT residing in the fact that MUST NOT mainly expresses strong prohibition was not confirmed. Strong prohibition as such seems rarely to be expressed at all, the main interpretation of MUST NOT being subjective moral impropriety. As far as the difference between the two styles is concerned, the journalistic style seems to prefer 'more objective means', both in the meaning a certain modal expression is used with as well as the choice of the means themselves.

Keywords: parallel corpus, comparative study, negative necessity, fiction, journalese

Introduction

The idea for a study based on a corpus resulted from the conclusions of my dissertation thesis whose aim was to compare means used to express necessity in English and in Czech based on contemporary fiction and the corresponding published translations. It has been discovered that the difference in meaning between CANNOT and MUST NOT is, by no means, clear and that the generally assumed correlation between English MUST NOT and Czech NESMĚT, and English CANNOT and Czech NEMOCI is not always valid. The scope of the dissertation, however, did not provide enough material to allow the deduction of any general conclusions.

As a first step, two years ago, corpora were used to study the matter in the area of fiction. The main goal of that first step was to investigate the basic means used to express 'obligation not to' with the use of corpora in both English and Czech. First the relevant means from the Czech National Corpus and the British National Corpus were collected, then the means were classified according to their different interpretations based on concrete criteria (e.g. objective/subjective modality, a ban/moral impropriety, etc.), it was followed by the study of the reasons why the particular means were selected based on the study of co-text (collocations, grammatical, pragmatic and stylistic aspects. The process was concluded by

contrastive study of the choices of the corresponding means in the two languages based on the parallel corpus Intercorp.

The principal aim of this article is to compare the data acquired in the first step mentioned above and compare it with the data acquired in the same way, but this time from the parts of the respective corpora including texts of journalistic style.

Modality

Modality represents a universal concept. According to Palmer (1986: 7), "It is probable that there are very few languages that do not have some kind of grammatical system of modality." It represents an extensive and complex problem and since the fundamental means of its expression differ from language to language, the starting points of its description can vary between languages.

Modality in general can be understood as *relation*: the relation of the speaker to the utterance, to the factuality and actualisation, the relation of the utterance to reality, the relation of the real world to possible worlds. Modality can be best seen as opposed to factuality; it states the content of the utterance not as a fact but as a potential fact dependent on certain conditions, e.g. the authority and approach of the speaker (*You must take the exam now; you can take the exam now*), the reliability of their judgement (*He must have forgotten about the meeting*).

Generally, two basic concepts of modality are distinguished, those of possibility and necessity. Each of these two concepts, necessity and possibility, is then further classified into kinds of modality: deontic, epistemic and often also dynamic.

One additional feature of modality which also appears in many works dealing with modality is the strength of modality. The strengths are referred to by different names by different linguists, e.g. Huddleston and Pullum (2002: 175) talk about "weak", "medium" and "strong", while Halliday (1991: 182) refers to the same concepts as "high", "median" and "low" although they describe the same matter. On the basis of strength, the fundamental concepts of *obligation* and *high probability* on the "high" end of the scale and of *possibility* and *permission* on the "low" one are distinguished. Medium modality is somewhere in between, "though intuitively closer to the strong end than to the weak" (Huddleston and Pullum 2002: 177).

The basic distinction between deontic and epistemic modality is that between actualisation and factuality. Deontic modality has its illocutionary force and an utterance including deontic modality has the potential (another important key term connected with modality) to result in a particular human behaviour. An utterance including epistemic modality, on the other hand, has the potential of being/not being true. Halliday (1991: 183) describes deontic modality as calibrating the area of meaning between *Do it!* and *Don't do it!*, whereas epistemic modality as calibrating the area of meaning which lies between *Yes* and *No*.

An additional kind of modality can be distinguished - dynamic modality. For dynamic modality the key words are properties (of a situation) or disposition (of a person). Dynamic modality can be, therefore, seen as more objective than the two previous kinds. Since, however, a common practice is to distinguish only two basic kinds of modality, such an approach is also adopted in this paper.

The distinction of the area of modality into the two aforementioned kinds is not absolute. In many cases it is a question of gradience rather than of clearly cut boundaries. The kind of interpretation is usually dependent on context because many modal expressions can be used with both deontic and epistemic interpretations.

'Obligation not to' in English and Czech: Theoretical Foundations

Grammar books do not seem to deal in any big detail with differences in the meaning and use between the individual means. One difference discussed is the difference between MUST on the one hand and a set of negative modal means such as CANNOT, MAY NOT, BE NOT ALLOWED on the other hand, which is described as the difference between 'necessity not to do' and negative 'permission to do' = a ban. The former is felt stronger. This assumption is commented on by Leech (2004), and Palmer (1979): "Both these statements (with *may not* and *mustn't*) are prohibitions, but differ in that the second sounds rather more forceful, positively forbidding instead of negatively withholding permission" (Leech 2004: 95), "But there is an obvious difference between refusing permission (*may not/can't*) and laying an obligation not to (*mustn't*). With the former it is to be assumed that permission is normally required, while with the latter the speaker takes a positive step in preventing the action for which permission may not normally be required." (Palmer 1979: 65)

In Czech there are two basic modal verbs used: NESMĚT and NEMOCI which, since they are both used with external negation, express negative permission. The only formal possibility in Czech to express the meaning of MUSTN'T is MUSET with the negative infinitive which is rare and restricted. Based on a short comment in Dušková (1994: 192): "The usual negation to extrinsic *may* is *must not*, or, *can't*, for example *you mustn't worry*: *nesmíte si dělat starosti*, *we can't take the dog into the hotel*: *nemůžeme (nesmíme) vzít psa do hotelu*. *One mustn't be proud*: *Člověk nesmí být pyšný*." One might arrive at the conclusion that MUSTN'T corresponds to Czech NESMĚT and CAN'T to Czech NEMOCI. This might be further supported by looking into dictionaries which either do not deal with the issue at all or suggest the same correlation, e. g. in Lingea Lexicon NESMĚT = MUST NOT, MAY NOT. The situation, however, may not be so easy. Grammatically, CANNOT should be close in meaning to MAY NOT since they both take external negation. One clear distinction of MAY NOT from the other modal verbs used to express 'necessity not to' is the formal nature of MAY NOT. "Negation *may not* (in extrinsic meaning "*nesmět*") occurs in official style." (Dušková 1994: 191)

As reference materials did not provide any further information, native speakers of British English were asked for help. The opinions they provided were interesting, and surprising at the same time. They agreed on the assumption that while there exists a rather big overlap between MUST NOT and CANNOT, partly depending on a particular context and intonation, there is a basic difference felt between the two means, which does not seem primarily to reside in the force of the modal means as Leech (2004) suggested, but rather in the meaning. While CANNOT expresses either inability (which is outside the scope of this paper) or something which is officially (often by law) not allowed/banned, MUST NOT is used to express more personal involvement of the speaker describing something which is considered morally wrong. It can be seen as more or less in agreement with Palmer (1971) (see the paragraph above). If something is usually not allowed, then one, if they in spite of this want or need to do it, has to ask for permission. It is, however, not logical to ask for permission to do something which is considered morally inappropriate. Such things are usually done either out of ignorance, or because one cannot help it.

Reference material on 'necessity not to' in Czech does not offer much information either. It was, therefore, attempted to deduce the negative meanings of the two modal verbs (NESMĚT, NEMOCI) from the positive ones. From Benešova's (1971: 132) classification of modal verbs it follows that MOCI is used to express possibility in general - it means in cases when the source of modality are outer circumstances as well as if the source of modality is a concrete human being and both in the case when the source of modality is identical with the source of action and when the source of modality is different from the source of action.

SMĚT is used only if the source of modality is a concrete person and only in cases when the source of modality is different from the source of action. Based on these findings, NESMĚT seems mainly to express things which are not allowed or/and which are officially banned. NEMOCI, on the other hand, refers, to inability and impossibility based on outer circumstances, to things seen as bad, something that a kind of inner control prevents a person from doing, e.g. *to jí nemohu udělat, nemohu ji zradit*. According to this explanation, therefore, the correspondence between MUST NOT, CAN NOT on the one hand, and NESMĚT, NEMOCI on the other should be completely opposite. An example from Rowling's Harry Potter might be used to support this view:

They cannot keep the objects longer than that unless they can prove they are dangerous. (= it is banned by the Ministry) - *Nesmějí si věci z pozůstalosti ponechávat déle, neprokáží-li...*

Since, however, in the same way as in English, also in Czech there is a big overlap between the two means, there is certain flexibility in their use.

The comparison becomes simpler when the discussed modal verbs are used with past reference because due to formal reasons MUST NOT cannot be used. MUST does not have the past form and in order to express strong obligation with past reference other forms must be used, the most common being HAVE TO. The negated form of HAVE TO, however, does not express an *obligation not to*, but a *possibility not to or the lack of obligation*, which hold true both in the present and past tenses. Apart from the just mentioned formal reason, there might be a semantic or logical reason for the non-existence of the past form of MUST with deontic interpretation. Since MUST, and in the same way probably also MUST NOT, is strongly subjective and being used when the source of modality is the speaker, it is highly illogical with past reference (even more so in its negative form). MUST NOT expresses something seen as wrong by the speaker, so if the speaker is at the same time the source of the action, the logical result has to be the non-realization of the action (no space for MUSTN'T here) or the action is accidental and then other means is preferable, e.g. *I did not want, mean, intend....* It is also possible to speak about a ban in the past and then COULD NOT, BE NOT ALLOWED, BE NOT TO etc. is used.

Data Collection

'Necessity not to' was studied on three corpora: the Czech National Corpus (CNK), the British National Corpus (BNC) and the parallel corpus Intercorp. Since the British National Corpus includes texts mainly from the 1990s, in Czech National Corpus part SYN2000 was chosen as the corresponding source. The corpus Intercorp, since it is comparatively smaller, was not reduced as far as dates of publishing of the texts are concerned, since the issue studied is not likely to be significantly influenced by the time difference in the order of decades. However, only works of fiction were selected from the "core part" of the corpus and then for the second step only news articles were selected from the so-called "collection" part. The same filters (fiction, journalese) were applied to the other two corpora. The individual means studied were selected on the basis of the above mentioned dissertation. For English the following means were included: MUST NOT, CANNOT, COULD NOT, MAY NOT, NOT ALLOWED TO, NOT SUPPOSED TO, BE NOT TO. For Czech NESMĚT and NEMOCI were studied. In the case of NEMOCI and negative forms of CAN and MAY which, apart from 'necessity not to' express also other meanings (which are in all the cases considerably more frequent), only cases where these means clearly expressed 'strong obligation not to' were included.

In the cases when the corpora provided a large number of tokens, 250 examples were processed for each style. In the cases when less than 250 examples were available or

manageable, always at least 50 relevant cases were included in the study, with the exception of MAY NOT for which from 411 cases only 19 were relevant in the area of fiction and from 560 cases in the area of journalism 67 were relevant.

The whole issue was first studied for each language separately. The means found in the corpora were classified into categories on the following criteria: the source of modality and the source of action (on this basis modality is classified into subjective or objective) and within each category further two categories were distinguished: moral impropriety and a ban.

Results of the Corpus-Based Study

The results gathered in the way described above are first presented in tables and then commented on. Tables 1 and 3 present data gathered from BNC, Tables 2 and 4 from CNK. The first pair of tables include the data collected from fiction and the second pair include texts of journalistic style.

Tables 1 and 2

Classification of the Means from the BNC and the CNK - fiction

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral improper (%)	Ban (%)	Moral impropr. (%)	Ban (%)
MUST NOT	71	12	7	10
CANNOT	57		17	26
COULD NOT	78			22
MAY NOT		60		30
NOT ALLOWED				100
BE NOT TO	11	11	12	66
NOT SUPPOSED	8		44	48

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral impropr. (%)	Ban (%)	Moral impropr. (%)	Ban (in %)
NESMET	34	19		47
NEMOCI	28		41	31

Tables 3 and 4

Classification of the Means from the BNC and the CNK – journalese

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral improper (%)	Ban (%)	Moral impropr. (%)	Ban (%)
MUST NOT	47	5	16	32
CANNOT	39		16	45
COULD NOT	75		6	19
MAY NOT		24		76
NOT ALLOWED				100
BE NOT TO		3	8	89
NOT SUPPOSED	6		42	52

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral impropr. (%)	Ban (%)	Moral impropr. (%)	Ban (in %)
NESMET	12	14	9	65
NEMOCI	32	21	22	25

The data from the BNC confirm that MUST NOT expresses mainly subjective modality, which is a generally accepted truth, and they also confirm the assumption of the native speakers discussed above that MUST NOT is predominantly connected with activities seen as morally inappropriate. What is rather surprising are the findings connected with the modal verb CAN. In fiction in both present and past forms it mainly expresses subjective modality and moral impropriety. The higher percentage of moral impropriety meanings connected with the past form of CAN can be explained by the fact that since MUST NOT, which is supposed to be a chief means to express moral impropriety, does not have a past form and

COULD NOT functions here as a suppletive form. This also applies to COULD NOT in journalese. In the present form, however, it is slightly more objective.

Another interesting fact which can be inferred from the data is the finding that for a ban English often uses other means than modal verbs. NOT ALLOWED TO, NOT SUPPOSED TO and NOT BE TO all predominantly express objective modality and a ban, both in fiction and in journalese where it is even more visible.

The relative closeness of meanings of MUST NOT and CANNOT/COULD NOT when expressing 'necessity not to' is further confirmed when the right-hand collocates of the two forms are studied. Among the 30 most frequent right-hand collocates of MUST NOT 16 were also among the most frequent collocates for CAN/COULD NOT. These were verbs *let, refuse, blame, complain, tell, say, get, risk, leave, allow, accept, pretend, disagree, ignore*. All these activities can rather be seen as morally inappropriate since they are not generally permitted or banned.

When fiction and journalese are compared, generally there are not many differences, as has already been pointed out when discussing some particular cases above. Nevertheless, a general tendency towards objective modality can be observed which is most visible in the modal expression MAY NOT, but can also be illustrated in the following ones: MUST NOT, BE NOT TO, NOT SUPPOSED TO. This can be caused by the fact that journalistic style as a whole inherently tends to be more objective than fiction. As regards MAY NOT, there was a difference in the proportion in which it appeared with the meaning of 'obligation not to' which was more than three times bigger in journalese than in fiction.

The issue was then studied on the parallel corpus Intercorp. English forms served as the starting forms in the search and their corresponding Czech translations were excerpted as well. The results are summarized in Tables 5 and 6.

Table 5 and 6

Means from Inrecorp – fiction and journalese

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral improper (%)	Ban (%)	Moral impropr. (%)	Ban (%)
MUST NOT	nesmět – 74, no C (no explicit modal means in Czech) - 5 nemoci - 38		nesmět - 21	
CANNOT	nesmět - 6 nelze upřít - 6 nemoci - 60		nesmět - 13	nesmět – 16 nemoci - 27
COULD NOT	nesmět – 12 no C - 8	Nesmět - 2 nemoci - 1	nemoci - 3 nesmět - 1 zakázat - 1	nemoci - 9 nesmět - 7
NOT ALLOWED				nesmět - 80 nedovolit – 20
MAY NOT				nesmět - 100
NOT SUPPOSED	nemít - 16	nemít - 21	nemít - 10 nesmět – 11 no C - 5	nemít - 21 nesmět - 16
BE NOT TO	nemoci – 7	nedovolit - 7 imperrative - 7	nemít - 7	mít zakázáno - 43 nesmět – 29

Modal means	SUBJECTIVE		OBJECTIVE	
	Moral improper (%)	Ban (%)	Moral impropr. (%)	Ban (%)
MUST NOT	nesmět – 32, no C (no explicit modal means in Czech) - 8	nesmět - 12	nesmět - 18	nesmět - 32
CANNOT	nemoci - 17 nesmět - 6	nemoci - 21	nesmět - 13	nesmět – 16 nemoci - 27
COULD NOT	nemoci - 48 nesmět – 12 no C - 10	nemoci - 1	nemoci - 8 nesmět - 2 zakázat - 8	nemoci - 9 nesmět - 2
NOT ALLOWED				nesmět - 64 nedovolit – 28 zakázat - 8
MAY NOT				nesmět - 100
NOT SUPPOSED	Nemít - 20	nemít - 22	nemít - 20 nesmět – 3, no C - 5	nemít - 21 nesmět - 9
BE NOT TO	nemoci – 4 nemít - 6	nedovolit – 5 imperrative - 1	nemít - 3	mít zakázáno - 43 nesmět – 18 nemoci - 20

In the distribution of basic interpretations of English means expressing ‘necessity not to’ the data from the Intercorp correspond for both the styles to those gained from BNC. As far as their translations into Czech are concerned, MUST NOT is mainly translated as NESMĚT. NESMĚT is, however, also the most frequent translation for CAN NOT. The data from Intercorp show that objective ‘necessity not to’ is in English frequently expressed by other means than modal verbs. This claim is further supported by an additional search in which English means used to translate Czech NESMĚT were excerpted with the following results: imperative construction, infinitive construction, WOULD NOT, NOT WANT TO, no C, NOT PERMIT (the means are ordered according to their frequencies which were in the range from 17 to 6 percent). When fiction and journalese are compared, more significant differences can be seen in the following areas. In general, NEMOCI is more frequent in journalese, used also in contexts where in fiction NESMĚT would be more likely to be found. The general choice of means seems more objective (objective moral impropriety or a ban) as was also the case of the two general corpora. All corpora also suggest a close correspondence between MUST NOT and NEMOCI since both these modal means are frequently used to express moral impropriety. This finding is further supported by the study of right-hand collocates of NEMOCI where the means expressing activities which are unlikely to be banned, such as *tvrdit, nechat, zklamat, říci, upírat, přesvědčovat, souhlasit, žádat, kárat* etc. prevail.

Conclusion

Corpora enable to study certain features of language on a large amount of material which was not possible or extremely time consuming in the past, and thus to distinguish delicate shades of meaning. The study presented here is an attempt at such a delimitation of different meanings of means used to express ‘necessity not to’. The generally assumed distinction between MUST NOT and CAN NOT residing in the fact that MUST NOT mainly expresses strong prohibition was not confirmed. Strong prohibition as such seems rarely to be expressed at all, the main interpretation of MUST NOT being subjective moral impropriety. The assumption that CANNOT, unlike MUST NOT, often expresses activities which require permission and therefore can be banned is valid only to a certain extent since the predominant meaning of CANNOT is also to express moral impropriety, although, in contradiction to MUST NOT, it expresses objective modality more frequently.

In terms of the difference between fiction and journalese, in the area of ‘obligation not to’, no significant differences were observed, apart from a general tendency of the journalistic style to prefer ‘more objective means’, both in the meaning that a certain modal expression is used with as well as the choice of the means themselves (e.g. more frequent use of NEMOCI observed both in CNK and Intercorp).

Non-native users of both English and Czech should be aware of possible oversimplification of the correspondence between MUST NOT and NESMĚT on the one hand and CANNOT and NEMOCI on the other and choose the corresponding means carefully and from a larger set of means available.

References

- Benešová, E., Panevová, J. and Sgall, P. 1971. *Čas a modalita v češtině*, Praha: Univerzita Karlova
- Čermák, F. and Blatná, R. 2005. Jak využívat Český národní korpus. Praha: Vivas.
- Dušková, L. et al. 1994. Mluvnice současné angličtiny na pozadí češtiny. Praha: Academia.
- Halliday, M. A. K. 1991. *An Introduction to Functional Grammar*. London: Arnold.
- Huddleston, R. and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press
- Leech, G. 2004. *Meaning and the English Verb*. London: Longman.
- Palmer, F., R. 1979. *Modality and the English Verb*. London: Longman.
- Palmer, F., R. 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
- Palmer, F., R. 1995. Negation and the modals of possibility and necessity, in *Modality in Grammar and Discourse*, ed. Bybee J. and S. Fleischman, Amsterdam/Philadelphia: John Benjamins, 454-471.
- Šimůnková, R. (2009) Analysis of the Means Expressing Strong “Necessity Not To” in English and Czech based on General and Parallel Corpora. In J. Levická (Ed.) *NLP, Corpus Linguistics, Corpus Based Grammar Research*, 349–35. Tribun.

Research Sources

- Czech National Corpus: SYN 2000. Prague, 2000. Available on-line <https://ucnk.ff.cuni.cz/>
- Czech National Corpus: Intercorp. Prague, 2006. Available on-line from <https://ucnk.ff.cuni.cz/intercorp/?req=page:info>
- British National Corpus. XML edition. Oxford, 2007