English and Russian vague category markers in business discourse: Linguistic identity aspects

Elena Malyuga*, Michael McCarthy
RUDN University, Russia

ARTICLE INFO

Article history:
Received 29 January 2018
Received in revised form 21 July 2018
Accepted 21 July 2018

Keywords:
Vague category marker
VCM
Business discourse
Corpus analysis
CANBEC
Russian National Corpus

ABSTRACT

Vague category markers (hereafter VCMs), also known as general extenders, are a pervasive phenomenon of spoken discourse. They include expressions such as and things like that and or whatever. They have been studied in conversational contexts and specialised contexts (e.g. courtroom discourse, radio broadcasts) but spoken business and professional communication has received relatively less attention. Using two corpora, this article addresses: (1) the forms and functions of VCMs in English business talk and in Russian business/professional talk, and (2) the comparability of VCMs across the two datasets. In both corpora, a range of VCMs similar to those found in everyday conversational contexts occur. The functions of VCMs in business/professional data replicate those illustrated in previous research into VCM use, i.e., the projection of fluid, exemplar-based categories which appeal to shared knowledge, hedging, the projection of a shared identity both within and between groups and as shorthand references to different levels of shared knowledge, from internal knowledge shared by the group to general, global knowledge and experience. The efficient functioning of VCMs is evidenced in turn-taking. VCMs in both datasets attach to a wide range of exemplar-types, regardless of syntactic configuration. Although the two datasets could not be perfectly matched, sufficient similarities enable useful comparisons to be made, albeit translatability of VCMs is often complicated by the number of internal variants any VCM may display.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

This article is concerned with the use of Vague Category Markers (VCMs, also known as General Extenders) in business and professional contexts in English and Russian. Both languages are well-researched, and both are blessed with sources of data for the study of business and professional discourse. The two languages, as may be assumed for many other languages, have developed conventions for the conduct in situations such as business meetings, negotiations, professional gatherings and so on. The topic which we wish to address here is: to what extent is vague category marking similar or different in English and Russian business discourse and what implications does such an investigation have for translation and cross-cultural understanding? The topic is essentially a socio-pragmatic one, concerned with the creation and interpretation of meanings in circumscribed contexts. Our focus will be on two sub-themes:
(1) the occurrence of VCMs in English business discourse and their comparability/translatability with regard to Russian; 
(2) the relationship between vague category marking and turn-taking in spoken business communication.

These foci are chosen since at the heart of our argument is the contention that vague category marking in discourse involves a joint enterprise between speaker and listener to achieve its pragmatic goals.

We use two corpora of business/professional data in the present article: the Cambridge and Nottingham Business English Corpus (hereafter CANBEC) and a sub-corpus of the Russian National Corpus (hereafter the RNC), both of which are described in Section 3.

2. Vague category marking

2.1. General outline and definition

Vague category marking refers to the purposeful use of a class of lexi-co-grammatical strings which make non-explicit references to categories which the speaker/writer assumes the listener/reader will be able to interpret without the need for the sender to give an exhaustive list of possible items in any individual category (or indeed, the sender may not have the requisite knowledge to be able to give explicit references or may hesitate to do so for a variety of reasons). Extract 1 below, from an English business meeting where a new method of processing customer orders is being discussed, suggests that the speaker does not feel the need to elaborate a list of product lines and can assume his interlocutors will understand the potential scope of and all that sort of thing:

Extract 1

Any teams which have ongoing lines like toothpaste and all that that sort of thing, they will work on the new way... (CANBEC)

The string that sort/kind of thing (sometimes preceded by and and/or all) occurs 52 times in the spoken English business data. The core string and its variants represent classic VCMs. Other common English VCMs include (and) things like that and or whatever.

The normality of the use of VCMs in spoken discourse and their role in the successful progression of speaking turns has been noted by, among others, Channell (1994), Cutting (2007), Pichler and Levey (2010), Sabet and Zhang (2016) and Vaughan et al. (2017). The term we use here, VCM, also occurs in the literature as extension particle, set marking tag, general extender, generalised list completer, vague tag and vague category identifier, among other labels for the phenomenon (see Dines, 1980; Jefferson, 1990; Dubois, 1992; Channell, 1994; Overstreet, 1999, 2005; Overstreet and Yule, 1997; O’Keeffe, 2004; Cheshire, 2007). A syntactic distinction is often drawn between adjunctive VCMs (those introduced by and) and disjunctive VCMs (those introduced by or, as in or whatever). The distinction comes from the work of Overstreet (1999:3–4), who correlates the two types with distinct pragmatic functions relating to politeness strategies, while Aijmer (2015) suggests that the disjunctive types may offer listeners the possibility of alternative interpretations of referents of the VCM.

For our purposes, we define a VCM as a phrase attached to an item or items seen as an exemplar or exemplars of a category of people, things, states or actions; the VCM extends the exemplar(s) in a non-specific way and typically encodes an assumption that people in the discourse will understand what might be included in the category or at least not feel the need to challenge it or demand an elaboration of its meaning. Thus, in extract 1, toothpaste is the exemplar chosen to represent the category of products under discussion; the VCM phrase is and all that sort of thing, which projects a category that includes similar items. The vagueness of the VCM serves two main purposes: (1) to project an assumption of shared knowledge or common ground – “intersubjectivity” as Overstreet (1999:293) terms it; and (2) to refer to categories which are valid for the participants at the time of utterance without any implications as to the permanence or universality of the category or to its being a closed lexical set (Overstreet, 1999:297; Cheshire, 2007:163).

Vague category marking often encodes what has been termed “instantial hyponymy” (Carter and McCarthy, 1988), i.e., exemplars are (co-)hyponyms of superordinates which are of the moment, which are fluid and open to the negotiation of meaning. However, it is rare in spoken contexts, though not unknown, in circumstances like those in extract 1, for any listener to challenge the VCM. Overwhelmingly, listeners appear, on the face of it, to understand what is to be included (Jucker et al., 2003). Vague category marking is simply heard as a normal, banal practice and it is evident that speakers and writers do it with “audience design” in mind (Bell, 1984, albeit Bell was concerned with matters of style). Unlike the radio audiences in Bell’s research, the business and professional contexts studied for the present paper include just two types of recipient: addressees (who could be individuals, or everyone present) and other, ratified recipients who may not be directly addressed, and in business and professional contexts, these two types of recipient are in general assumed to be in the know with regard to the context.

VCMs refer to categories but do so by deliberately de-focusing from available lexicalised categories (e.g. taxonomies such as plants, mammals, vehicles) and creating ad hoc, instantial categories relevant to the needs of the present discourse. The scope of their reference is projected as sufficiently interpretable to the recipient(s) and need have no permanent validity. Toothpaste and all that that sort of thing needs only to be valid for the understanding of product lines at the time of speaking among that group of participants and their shared knowledge of what the company manufactures. There is no implication
that every conceivable item of personal hygiene is being referred to. The category boundaries lie within the shared knowledge of those who work in the company in relation to what product lines may be affected at that time.

This raises the question of what kinds of knowledge interlocutors draw upon to use and interpret VCMs, a question put succinctly by Overstreet (1999:18): “How interlocutors manage to create a compatibility of categories, despite the fact that individual mental worlds are necessarily distinct”. In a study of the occurrence of VCMs in a radio phone-in programme broadcast in the Republic of Ireland, O’Keeffe (2003, 2006) showed that distinct levels of knowledge were projected by the VCMs occurring in her data. Some of these referred to domains which were probably only interpretable to Irish listeners, some operated on a wider, European scale and some were global (i.e. interpretable by almost anyone). More particularly, O’Keeffe (2003) related the use of VCMs, along with other features such as pronoun choice, to the creation of identities. The VCMs in her data which referred to Irish cultural and national matters enabled the cohort of listeners to identify as a social and cultural group sharing a concern with Irish current affairs. We may hypothesise from this that similar distinct levels of shared knowledge may operate in Russian and English in the business context, with global knowledge, national and regional knowledge, knowledge of the business and professional world, knowledge of particular products and processes, and knowledge of the products and processes of an individual company all playing a part in the interpretability of the VCMs in the data. Equally, we may suppose that the creation and maintenance of identities (whether corporate, national or cultural) is supported by VCM use.

VCMs as part of in-group language practices provide substantial grounds for the analysis of linguistic identity, because the way VCMs are interpreted very much depends on the way conversational participants identify themselves in terms of the language they produce and project. Projected shared knowledge serves as a cornerstone for adequate understanding of VCMs by all those involved in the interaction, for it not only unites the participants as a group of insiders capable of interpreting the meaning potential of the VCMs, but also singles out each individual member of this group by identifying him or her as a bearer of shared knowledge. This is particularly true of business and professional communities, where shared identity and shared goals are typically paramount and where the successful persecution of activities depends greatly on the assumption that one’s interlocutors, both within and between groups, have in common a basic perspective, prior experience and/or training and a shared linguistic repertoire.

Because business/professional discourse serves as a common denominator for the shaping of knowledge and action shared by those involved in its production (Malyuga, 2018), its various elements (VCMs included) form a shared linguistic code that at all times relies on mutual understanding, or at least a desire for such. This idea underlines the somewhat “closed” nature of VCMs used in any one particular discourse, for example a business meeting or a sales negotiation, as it implies marking of not only linguistic categories, but also of the speakers involved in their production and perception, for this is where a person will be labelled either an insider or an outsider within the discourse community. For this reason, Handford’s (2010) suggestion of the “tactical” role of VCMs in business meetings with participants from outside of the host company, where the VCM projects a bonded in-group even before such a group may gel in reality, accords well with notions of identity creation and reinforcement.

2.2. VCMs in different settings

The primary locus for the study of VCMs has been informal conversation among native speakers, a fact not surprising given the long acceptance of conversational data as a benchmark for the investigation of other kinds of spoken data (Sacks et al., 1974; Drew and Heritage, 1992). Other types of data such as meetings, service encounters, health communication and academic discourse are often interpreted in terms of their closeness to or diversion from casual conversation. Informal conversational settings for the investigation of VCMs are included in works by Overstreet (1999), Jucker et al. (2003), Evison et al. (2007), Terraschke and Holmes (2007).

Non-conversational settings for the study of VCMs include Cotterill’s (2007) study of courtroom discourse, O’Keeffe’s (2006) study of a radio phone-in programme and Adolphs et al. (2007) in the field of health communication. Studies have also focused on language learners’ use of vague language, with some evidence of under-use of VCMs in the target language when compared with native-speaker use of VCMs, though the differences may not be great (Xu, 2016; Lin, 2013). Business settings have been less explored, though Liu (2015) examines the use of vague language in cross-cultural written business communication and sees vagueness as related to politeness, protection of face and as a strategy for achieving business goals, which aligns with some of our present findings.

Cross-cultural contexts have been explored in Cucchi’s (2007) study of vague language in Italian and English in the context of the European Parliament, which notes the differing degrees of power and distance and familiarity encoded in the use of VCMs in the two languages. Cheng and Warren (2001), who recorded the use of vague language among native and non-native speakers of English in Hong Kong, found no evidence that differences in vague language use caused problems of cross-cultural understanding, with speakers in both datasets using vague language to create informality and mutual accommodation. An investigation of the frequency and grammatical distribution of VCMs in Persian by Parvaresh et al., (2010) concluded that adjunctive VCMs were more frequently used than disjunctive VCMs (see also Fernandez and Yuldashev, 2011), but for Dutch versus English usage among learners in Buyssse’s (2014) study, disjunctive VCMs were more common in the Dutch data. Combining Persian, Chinese and English data, Sabet and Zhang (2016) found that cultural, linguistic and, in the academic context of their study, pedagogical factors all played a role in differences in the use of vague language when L1 speakers of English were compared with Persian and Chinese counterparts. All these studies point to the highly context-dependent use of VCMs.
In relation to the question of possible translation equivalents of VCMs across languages, which the present article addresses, Terraschke and Holmes (2007:212) observe that apparently equivalent forms may have a different distribution and function across languages (in their case, German and English). Meanwhile, Fernandez (2013) looks at cross-linguistic questions of VCM use and occasional mistranslation by English-speaking learners of Spanish; Terraschke (2007) notes similar problems among German speakers of English. Ruzaitė (2010) investigates vague language (including VCMs) in a parallel corpus of Lithuanian and English written data and also raises questions of translatability, noting how apparently equivalent VCMs across the two languages can vary in their frequency and how a bilingual dictionary may fail to reflect actual usage.

2.3. The functions of VCMs

Much discussion has taken place on the pragmatic functions of VCMs. Their purposeful use has been seen in terms of hedging, either out of genuine lack of precise or accurate knowledge on the part of the speaker or as a politeness strategy to forestall threats to face (Overstreet, 1999; Jucker et al., 2003:1765; Koester, 2010:156–158). Koester (op. cit.:62) notes the use of VCMs to refer implicitly to facts and information while engaged in explanations. Overstreet (1999) also refers to attitudinal uses to create intensification and reactions to previous elements in the discourse (see also Aijmer, 2015). VCM use has also been associated with creating solidarity among conversational participants and the creation of intersubjective meanings (Overstreet and Yule, 1997; Overstreet, 1999:99–104; Cutting, 2000, 2001). Handford (2010:164), who found that VCMs were (to his mind, counter-intuitively) more frequent in business meetings between companies and external parties than in internal company meetings, suggests that VCMs were used tactically to project an impression of shared knowledge rather than to rely on existing knowledge, thus creating a stronger bond and greater convergence between the parties who had not (yet) necessarily entered into any contractual arrangement. This matters in relation to the question of identity creation and reinforcement, in that group identity can be ‘projected’ even before it exists.

From their forms out of context, little can be said about VCMs in relation to these possible functions, nor can the utterance of a single speaker tell us much. Overstreet (1999) refers to the problem of interpretation and attempts to resolve it by citing data where the speakers are people known to her, with whose worlds she is familiar with. We enjoy no such luxury in the present paper, as large corpora collected from disparate sources can rarely offer more than happenstance familiarity with the data where the speakers are people known to her, with whose worlds she is familiar with. We enjoy no such luxury in the present paper, as large corpora collected from disparate sources can rarely offer more than happenstance familiarity with the speakers. In this paper, we argue that some of the key contextual factors are embedded in the turn-taking system. We argue that VCMs are more readily interpreted in terms of what response (if any) they elicit (Vaughan et al., 2017). This co-constructional approach to VCMs aligns with similar notions discussed in (Jucker et al., 2003) and (Cheng, 2007) and offers us at least some surface evidence of group activity and identity-forging.

3. Data and analytical approach

3.1. Data comparability

A recurring problem with corpus-based investigations is that of comparability. Especially in the case of informal spoken data, very rarely if ever will two corpora be perfectly balanced and equally sourced from different contexts of occurrence. Even recent, high-technology-based studies admit a lack of agreement over how one can exactly establish comparability (e.g. Li et al., 2018). Parallel corpora, consisting of translations of one set of documents into another to provide two datasets, offer one solution (e.g. Johansson and Hofland, 1994) but creating parallel corpora of spontaneous spoken language is a largely unfeasible task.

One possible way around the problem is to simply take the biggest possible available corpora in the hope that massive amounts of statistical data will yield large-scale generalisations and obscure local differences. In the present case, this could be achieved by, for example, comparing the British National Corpus (BNC) with the Russian National Corpus (RNC). However, even these two large corpora are hardly comparable, with the BNC being a collection of contemporary texts (see http://www.natcorp.ox.ac.uk/corpus/), while the RNC (see below) is a diachronic collection. No doubt generalisations at some level about the distribution of phenomena such as VCMs could be derived from a comparison of these two corpora, but they would tell us less about the contextual sensitivity of their use than smaller, contextually-targeted corpora, especially in relation to the creation and maintenance of group identities. The dilemma is thus to obtain as best a set of comparable data as is available, given the probability that no two corpora will have identical categorisation or annotation of their data. In the present case, the British data (see below) is more narrowly circumscribed than the Russian data, but together they provide business and professional contexts which, in our view, narrow down the types of speakers to a reasonably comparable, though imperfect level sufficiently to yield useful insights with regard to VCM distribution, translatability and functions within the broad world of business and professional life. In the next sections, we describe the data available for the present study.

However, because the two corpora were not fully comparable, instead of comparing two sets of statistical output based on identical analyses, the present article uses the English output as the basis for exploring the translatability of a selection of VCMs into Russian, looking at both semantic and pragmatic features. This was felt to serve one of our main purposes, the exploration of the straightforwardness or otherwise of translating VCMs and what the results of such an investigation might also reveal about the forging of business and professional group identities in the two languages.
3.2. English data

The CANBEC corpus, which provides the English data for the present study, consists of just over 900 000 running words of various genres of spoken business discourse. The data were collected from 2001 onwards in 26 different companies, including smaller and larger industrial and service enterprises (with employees ranging from 50 000 to fewer than 10), predominantly sourced from middle- or upper-management UK speakers, with around 10% of the speakers being users of English as an L2. The activities and industries covered in CANBEC include industrial equipment (e.g. cranes and lifting gear), pharmaceuticals, service industries (e.g. hotel and pub chains, financial services, consultancy and IT). Meetings were sometimes scheduled, had differing numbers of speakers and included external (inter-company) meetings and internal (intra-company) gatherings in a ratio of approximately 1:2.7. Topics included everyday issues and procedures, production schedules, decision-making, logistics, pricing, sales and marketing, human resource management and so on. The corpus was also designed to include what might be termed “academic business discourse” (e.g. lectures given at a university business school). Further details of the composition of the corpus and extensive analyses of its data may be found in Handford (2010).

3.3. Russian data

The Russian data come from the Russian National Corpus (RNC). The RNC incorporates both complete authentic texts and specific extracts, which can be helpful in tracing and interpreting items and contexts that could otherwise end up overlooked. The texts in the RNC are diverse in terms of age and gender groups, social and professional identities of speakers, the time and the geography of recordings, and cover a significant period of over 70 years, which enables scholars to trace changes taking place in the realm of spoken language, register the emergence of new trends, etc. Importantly, especially in the context of the present study, the RNC comprises transcripts covering various spheres of communication, including business and professional communication, and generated in various contexts (dialogues, oral reports, interviews, meetings, etc.). The RNC incorporates over 364 million tokens and currently offers the most representative corpus-based data for researchers investigating the Russian language (Russian National Corpus, 2018).

The RNC provides metatextual markup, whereby the texts can be filtered to meet specific research objectives. The basic metatextual features include (1) the sphere of functioning (public, non-public, cinema); (2) text type (conversation, interview, micro-dialogue, etc.); (3) text topic (private life, healthcare, politics and social life, etc.); (4) time of recording; (5) place of recording; (6) text style (neutral, colloquial, official); (7) audience characteristics (size, age, qualification).

The texts are categorised into a number of sub-corpora (Table 1).

<table>
<thead>
<tr>
<th>Sub-corpus</th>
<th>Number of texts</th>
<th>Number of tokens</th>
<th>Percentage of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>76 882</td>
<td>209 198 275</td>
<td>57.3%</td>
</tr>
<tr>
<td>Newspaper</td>
<td>181 175</td>
<td>113 292 003</td>
<td>31.0%</td>
</tr>
<tr>
<td>Dialectal</td>
<td>197</td>
<td>194 283</td>
<td>0.1%</td>
</tr>
<tr>
<td>Educational</td>
<td>229</td>
<td>664 751</td>
<td>0.2%</td>
</tr>
<tr>
<td>Parallel</td>
<td>370</td>
<td>24 022 437</td>
<td>6.6%</td>
</tr>
<tr>
<td>Poetry</td>
<td>41 448</td>
<td>6 738 474</td>
<td>1.8%</td>
</tr>
<tr>
<td>Multimedia</td>
<td>31 741</td>
<td>648 576</td>
<td>0.2%</td>
</tr>
<tr>
<td>Spoken Russian</td>
<td>3 034</td>
<td>10 122 579</td>
<td>2.8%</td>
</tr>
<tr>
<td>Spoken Russian: Business and professional</td>
<td>621</td>
<td>3 389 255</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Thus, the Business and Professional Language section of Spoken Russian sub-corpus consists of 621 texts and 3 389 255 tokens. The sub-corpus includes language material retrieved from business conversations, reports, interviews, conferences, round tables, lectures and meetings. Importantly, the RNC provides ample opportunities for the analysis of context, which is essential in investigating turn-taking phenomena.

Notably, one of the key disparities in terms of comparative analysis of the two corpora had to do with the difference in their organisation and content availability: while the CANBEC texts were analysed offline, the RNC only provides online versions of the material that cannot be retrieved for further analysis via offline corpus-analytical software. This variance in content presentation has motivated an inevitable disparity in methodologies applied to analyse the two datasets. Thus, the RNC online search engine does not offer the cluster function available in Wordsmith Tools and does not present the data indicating pause duration. This, however, did not put any substantial constraints on the procedure, since the data compiled using the RNC online search engine could be nonetheless compared with the CANBEC data against various indicators (frequency of occurrence, structural composition, contemporaneity, context and turn-taking).

With the RNC, the procedure for data retrieval essentially involved (1) switching to the “Spoken corpus” tab, (2) customising the subcorpus by selecting the required topic and text type (business conversations, reports, interviews, conferences, round tables, lectures and meetings), and (3) searching for the VCM while leaving the “Lexico-grammatical search” field parameters intact in order to ensure non-restricted data coverage.
4. Analysis and discussion

4.1. VCMs in CANBEC

The CANBEC data were analysed using the Wordsmith Tools software suite Version 7 (Scott, 2017). Searches were run on a list of VCMs identified by previous researchers in various contexts, principally from the work of one of the present authors and his co-researchers (see Evison et al., 2007), who identified 11 VCMs (including their variants) which occurred more than ten times in a 1-million-word sub-corpus of socialising and intimate conversations extracted from the 5-million-word CANCODE spoken corpus. These, along with comparisons of VCMs attested in other researchers’ data (e.g. Cotterill, 2007; Cucchi, 2010), provided a good range of search terms for the present study. Concordances were generated for each of the terms. Additionally, and to provide greater objectivity, the cluster function in Wordsmith Tools was used to compute clusters up to a maximum of six words, with a minimum occurrence of ten. The clusters generated were then checked manually against the selected search terms and read for the occurrence of further items (e.g. this, that and the other, which was not in the original search list, though this bore out its occurrence and functions discussed in McCarthy and Carter, 2002). VCMs occurring 10 times or more per million words are listed in Table 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Occurrences</th>
<th>Per 1 m words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>or something (like that)</td>
<td>201</td>
<td>223</td>
</tr>
<tr>
<td>2.</td>
<td>or whatever</td>
<td>174</td>
<td>193</td>
</tr>
<tr>
<td>3.</td>
<td>and whatever</td>
<td>162</td>
<td>180</td>
</tr>
<tr>
<td>4.</td>
<td>etc/et cetera/et cetera</td>
<td>130</td>
<td>144</td>
</tr>
<tr>
<td>5.</td>
<td>and everything (else)</td>
<td>88</td>
<td>98</td>
</tr>
<tr>
<td>6.</td>
<td>(and) (all) that/those sort(s)/type(s)/kind of thing</td>
<td>62</td>
<td>69</td>
</tr>
<tr>
<td>7.</td>
<td>(or) anything (like that)</td>
<td>60</td>
<td>67</td>
</tr>
<tr>
<td>8.</td>
<td>(and) stuff (like that)</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td>9.</td>
<td>and that</td>
<td>43</td>
<td>48</td>
</tr>
<tr>
<td>10.</td>
<td>and so on</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>11.</td>
<td>and this, that and the other</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>12.</td>
<td>(and) things like that</td>
<td>28</td>
<td>31</td>
</tr>
<tr>
<td>13.</td>
<td>and so forth</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>14.</td>
<td>(and) all that (sort/kind of) stuff</td>
<td>28</td>
<td>31</td>
</tr>
<tr>
<td>15.</td>
<td>and things</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>16.</td>
<td>and all that</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1163</td>
<td>1293</td>
</tr>
</tbody>
</table>

* Kind did not occur in the plural in a VCM.

A total of 16 items and their variants were found to occur 10 times or more per million words. After this, there was a cliff-edge drop-off in frequency, with remaining VCMs being considerably less frequent, for example, or whoever (six occurrences), or wherever (four occurrences), and all the rest of it and all sorts (two occurrences each), and all of that stuff, and suchlike, or stuff like that and or things like that (just one occurrence each). There are additionally five examples of what could be considered syntactically anomalous forms with these/those + singular sort/type/kind of + plural things. Thus, frequency varies considerably from well over a hundred down to one, with the two most frequent forms being disjunctive VCMs. Those two, along with item 7, mean that disjunctive VCMs account for some 37% of the top 16 VCMs.

4.2. Exemplars

The next stage of the analysis was to examine each occurrence of each VCM in the context of their concordance lines to ascertain the nature of the exemplars to which the VCMs were attached. This proved to be a not straightforward exercise in many cases, as cumulative items that were extended by a VCM often had to be traced back across turn boundaries and in some cases, the point where exemplars could be said to begin was impossible to pinpoint objectively. For example, in Extract 2, it is not clear whether the exemplar is the whole of the speaker turn, or and getting people to get back to me, or just to get back to me.

Extract 2

[<$\text{S1}$>, <$\text{S2}$>, etc. denote different speakers; <$\text{SE}$> 1 sec/0.5 sec <$\text{S1}$> denote pauses in seconds; − denotes aborted words; + denotes a latched utterance] <$\text{S2}$> So I’m still kind of chasing them up and <$\text{SE}$> 1 sec <$\text{S1}$> getting people to get back to me and stuff.

What is apparent is that this is a problem for the outsider-researcher and not for the participants, as we may observe when we add more context and note the turn-sequences (and, in this case, a second VCM); the listener’s two responses of right suggest there is no need for further elucidation.
Another example of the importance of turn-taking is when an interlocutor repeats the VCM (or part of it), which can be seen as evidence of convergence with the projected meaning or scope of the VCM, as in Extract 4:

Extract 4

<$2$> Obviously if we get leads <$E$> 1 sec < \ <$E$> <$=$> erm<$E$> 1.5 secs < \ <$E$> if the if < \ <$=$> we need to be wherever it is. We need to be in + <$1$> Mm. <$2$> +China in <$E$> 1 sec < \ <$E$> Korea or wherever + <$1$> Wherever. <$2$> +we need to be there. <$1$> That’s right.

Observing turn-taking phenomena, especially minimal and non-minimal response items (McCarthy, 2003), helps to support interpretation of the reception of VCMs, but considerable fuzziness remains for the analyst. However, that should not be a cause for concern, given the discussion in Section 2, above, of the instantial, implicit and ad hoc nature of the categories and assumptions or tactical projections of shared knowledge which VCMs encode at any given moment in the discourse and with the considerable fluidity immanent in category marking. The data clearly bear this out.

Exemplars may occur singly or as a list, as the following two extracts show.

Extract 5

<$1$> Yeah. But what I’m saying you need you need, all I’m saying is that you need to be away from here+
<$3$> Yeah.
<$1$> +in reasonable time+
<$2$> Yeah.
<$1$> + allowing for+
<$3$> Yeah.
<$1$> +allowing for traffic and everything else.
<$2$> Mhm.

Extract 6

<$6$> ... in this particular month there was a lot of work done whether it be fitting or modifying or extending or whatever which was beyond the standard.

An additional option, though less frequent, is to add another exemplar after the VCM, perhaps to further strengthen or clarify what is being projected. In this case, the speaker adds Halloween parties after the VCM and this that and the other.

Extract 7

<$7$> I know proper football clubs like Villa they do, but they spend a lot of money <$=$> on their <$E$> 1 sec < \ <$E$> on the er social side of it. They’re forever having dinner and dances and this that and the other and Halloween parties and you know get to meet your favourite footballers.
<$2$> Well a lot of them have clubs there now.
<$7$> Yeah.

Once again, in Extract 7, the evidence of turn-taking is important, with speaker 2’s contribution reinforcing the understanding of dinner, dances and parties having their typical context associated with the social clubs attached to many football teams.

One of the most interesting features of VCMs is their syntactic relationship to the exemplar(s). Considerable flexibility is again the case. Most straightforward are cases where a noun phrase or a series of noun phrases is extended by a noun-based VCM (i.e. one based around thing or stuff), as in utilities, councils, water companies, that sort of thing and that forces your channel intermediaries, your shops and things like that to stock the product, with the count nouns extended by syntactic matching with a count noun (whether singular or plural). Likewise, there are a number of straightforward cases of non-count nouns extended by non-count stuff, as in the bulk material and stuff. However, very many cases for all the VCMs analysed do not follow strict syntactic matching. Stuff, for example, is frequently used to extend count nouns (e.g. all the licences and stuff/the codes and stuff/your phone number and stuff), to extend verb phrases and clauses (e.g. to try and make sure that they were complying and stuff/pops out to the deli and stuff), Table 3 shows a sample of VCMs and the distribution of exemplar types.
Verb phrases and clauses are conflated in the table as these are often the most difficult types of exemplar to separate in terms of where the exemplar starts and finishes. If any insight emerges from the figures in Table 3, it is that wide variation in syntactic alignment is apparent. Adjectives seem the least favoured word-class for extension by these particular VCMs; and so forth seems to have a definite preference for extending verb phrases and clauses; and everything shows a preference for nouns, of all types. None of this should surprise us, since VCMs are, by their very nature, what are often referred to as chunks or fixed expressions, and, as such, have no “internal” grammar.

4.3. Functions

What we may conclude from the English data is that a wide range of VCMs occur in English spoken business communication and that the VCM types correspond closely with those found in everyday conversational settings, as evidenced in the literature previously cited. Their functions are, in the broadest sense, indices of projected shared perspectives. These perspectives are, as O’Keeffe (2003, 2006) argued, of different levels of generality. We have already seen an example in Section 2.1, above (toothpaste and all that sort of thing), where an adequate interpretation of the VCM depends on in-company knowledge of that company’s product lines. The example They’re forever having dinner and dances and this that and so forth and the other and Halloween parties, also quoted above, is a much broader type of projection of shared knowledge, yet still within a culture that would understand football teams and their associated social clubs, what a Halloween party is and what it might look like, and why it is a good exemplar of the kinds of social activities (in this case, alcoholic drinks) the company might cater for. At the same time, it projects a shared identity: its use reinforces membership of a society for which Halloween is a significant festival. Equally, many references would not be problematic for most people involved in business, such as Extract 6.

Extract 8

<$2$> And erm <$5$> pause <$\,$>$ our original customers were typically smaller less well-known customers <$5$> er g. <$\,$>$ <$5$> businesses.

<$1$> Mm.

<$2$> And a number of them were were dot coms and so on.

What it means to engage with dot coms and smaller businesses may be assumed to be within the experience and knowledge of a wide, global range of business people, not just in the UK or western context. Other references are very general and not confined to the business world, as in so it can be left probably until June, July, something like that. Having said that, the overwhelming majority of VCMs refer to business entities and activities, which, again, is hardly surprising, given the goal-orientation of business discourse and the need to create and reinforce cohorts with shared identities.

In business discourse, as Handford (2010) has argued, VCMs and vague expressions in general are important as tools in the delicate and hedged contexts of negotiation, especially in inter-company meetings, where Handford (ibid.) found a higher density of VCMs than in internal meetings. Handford argues that VCMs could be a tactic for projecting an assumed shared knowledge with those outside the company in order to create strong bonds and engagement. In the most basic sense, VCMs are a shorthand enabling users to avoid specification while opening the possibility of consideration of entities, states and activities not explicitly mentioned. The conceptual space opened up is ideal and fertile ground for the (often delicate) negotiation of agreements, decisions and joint perspectives and understandings. This is especially true of disjunctive VCMs, where the inclusion of or, for example, in a sales negotiation: And obviously you can tell us then how many you can produce a day or whatever to give us an idea, opens a wider door to conceptualisation, speculation and different possible courses of action than an adjective VCM might have done. In one meeting between two companies, the purpose of which is to review planning and progress, no fewer than 17 instances of or whatever are recorded, suggesting quite an open-ended review. Extract 7, from that meeting, has the interlocutor responding to the VCM with an utterance which is also somewhat vague (slightly different):
Having laid out the broad picture of VCM form and function in the English business data, we now turn to our Russian data to assess how, and to what extent, valid comparisons can be made across the two languages.

4.4. VCMs in the RNC

For the purposes of the present, initial comparison between the two languages, a set of six VCMs from the RNC were analysed. These represented (1) a disjunctive VCM, (2) a VCM based on a translation equivalent of thing, (3) a VCM based on a translation equivalent of stuff, (4) a VCM based on a translation equivalent of sort/type/kind and (6) a VCM based on a pronoun form (that). In this way, it was felt that the principal categories of the English VCMs were represented in the Russian searches as offering the basis for a broadly-grounded comparison. That being said, some translation equivalents based on this approach might be open for discussion: seeing that all VCMs share a common semantic load, translating them from any one language to another is likely to yield variable results. This fact, however, opens up promising opportunities for a further study of semantic and structural correlation between VCMs in various language pairs. The following direct translation correspondences to the English VCMs were arrived at:

1. в таком (этом) духе — or something (like that);
2. в таком (этом) роде — (and) things like that;
3. и все такое (прочее) — (and) stuff (like that);
4. и так далее — etc./etcetera/et cetera;
5. и тому подобное — (and) all that sort/type/kind of thing;
6. и прочее — and all that.

These search terms were based on a set of items outlined previously by Russian researchers (see Adamovich, 2011; Petrunina, 2015; Karasik, 2011).

Notably, though, translating Russian VCMs in search of their English equivalents does seem to yield rather subjective results, for the differences in their semantics between members of the selected set are far from straightforward. An obvious obstacle, for example, will be encountered when attempting to differentiate between the meanings expressed by VCMs like or something like that vs or anything like that, or etcetera vs and so forth. The same is true for Russian VCMs (cf. в таком духе vs в таком роде, etc.). However, for the purposes of this study, the above translation correspondences are presented as reasonably justified on the following grounds:

1. в таком (этом) духе — (and) things like that are given as translation equivalents because of structural (string length and composition) and closest semantic similarity (Russian дух in its contextual meaning in the given VCM can be readily traced to its English counterpart things based on their colloquial profile);
2. в таком (этом) роде — or something (like that) also present a structurally (for the same reasons as in (1)) and semantically similar set (the semantic content of Russian в роде is more suited to be paired with the English something, rather than its counterpart in things like that);
3. и все такое (прочее) — (and) stuff (like that) are also marked by prominently similar structural and semantic features;
4. и так далее — etc./etcetera/et cetera are listed as one of the top translation equivalents in dictionaries;
5. и тому подобное — (and) all that sort/type/kind of thing exhibit the closest correlation, as they are (a) both adjunctive VCMs, and (b) Russian подобный presents a direct translation equivalent for sort/type/kind;
6. и прочее — and all that are listed as one of the top translation equivalents in dictionaries, just as is the case with (4) above.

Following the same procedure as with the English VCMs, concordances were generated for each of the terms with the necessary omission of the clustering step, as this function, as mentioned, is not provided by the RNC search engine. VCMs occurring ten times or more per 1 million words are listed in Fig. 1.
Generally, four out of the six most frequent VCM phrases have two to 13 variants (see Table 4) which are primarily distinguished by introductory elements supplementing the core phrase. These elements include:

1. Conjunctions: и (and), или (or)
2. Indefinite pronouns: что-то, что-либо, что-нибудь (something)
3. Adverbs: где-нибудь (somewhere), еще (rendering the meaning something else of the kind)
4. Particle: вот (rendering the meaning that is).

In Table 4, the first figure in the “Occurrence ratio” column corresponds to occurrences of a particular VCM variant listed in the “Variations” column; the second figure corresponds to the total occurrences of the VCM in all of its variations.

Table 4
Structural variations and occurrence ratio of Russian VCMs (per 1 mln words).

<table>
<thead>
<tr>
<th>No.</th>
<th>VCM phrase</th>
<th>Variations</th>
<th>Occurrence ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>И все такое (прочее) (and) stuff (like that)</td>
<td>И все такое</td>
<td>51/66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>И все такое прочее</td>
<td>15/66</td>
</tr>
<tr>
<td>2</td>
<td>В таком (этом) духе (and) things like that</td>
<td>Что-то в таком духе</td>
<td>6/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Вот в таком духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>И в таком духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-нибудь в этом духе</td>
<td>2/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь в этом духе</td>
<td>4/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-то в этом духе</td>
<td>3/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в этом духе</td>
<td>9/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или еще что-то в этом духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в таком духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-нибудь в этом духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или че-го-то в этом духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Чем-то в этом духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>В таком духе</td>
<td>1/32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь в таком роде</td>
<td>2/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Где-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>И все в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-то в этом роде</td>
<td>6/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Чем-то в этом роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в этом роде</td>
<td>5/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь в этом роде</td>
<td>4/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или еще что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь еще в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td>3</td>
<td>В таком (этом) роде or something (like that)</td>
<td>Или что-нибудь в таком роде</td>
<td>2/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Где-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>И все в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-то в этом роде</td>
<td>6/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Чем-то в этом роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-то в этом роде</td>
<td>5/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь в этом роде</td>
<td>4/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или еще что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Что-то в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Или что-нибудь еще в таком роде</td>
<td>1/24</td>
</tr>
<tr>
<td>4</td>
<td>И так далее etc./etcetera/et cetera</td>
<td>И так далее</td>
<td>22/23</td>
</tr>
<tr>
<td>5</td>
<td>И прочее and all that</td>
<td>И так далее и тому подобное</td>
<td>1/23</td>
</tr>
<tr>
<td>6</td>
<td>И тому подобное (and) (all) that sort/type/kind of thing</td>
<td>И тому подобное</td>
<td>17/17</td>
</tr>
</tbody>
</table>
Notably, no occurrences were registered for Чё as a commonly expected spoken variation of ЧТО-НИБУДЬ, which is indicative of a yet persisting boundary separating the realms of business and day-to-day conversation in the Russian language.

4.5. Exemplars

The exemplars to which the Russian VCMs are attached fall into 8 categories, namely:

(1) CLAUSES, e.g.:
Extract 9
$<$S1$>$ У них в офисах карты были, инструкции, как управлять самолетом, и всё такое (They had maps in their offices, and instructions on how to fly a plane, and stuff like that).

(2) COUNTABLE NOUNS, e.g.:
Extract 10
$<$S1$>$ Перепродажи ведутся через аукцион, галерею изобразительного искусства, художественный салон, магазин и так далее (The reselling operations are executed via auctions, picture galleries, art shops, stores, etc.).

(3) UNCOUNTABLE NOUNS, e.g.:
Extract 11
$<$S1$>$ Если денег не будет, также все желание у бизнеса пропадет, начнется воровство и всё такое (If there’s no money, businesses will lose interest, we will see the onset of theft and stuff like that).

(4) PROPER NOUNS, e.g.:
Extract 12
$<$S1$>$ И опять же это влияние СМИ, когда вот эти “Пuppеты” и всё такое, потому что выставляют дураками всех (Yet again, there’s the influence of the media, with these “Puppets” and stuff like that, because everybody is made to look foolish).

(5) ADJECTIVES, e.g.:
Extract 13
$<$S1$>$ Значит, и очень простые, и очень банальные, или что-то в этом духе (Meaning very simple and very common, and things like that).

(6) VERBS, e.g.:
Extract 14
$<$S1$>$ Действительно, там химические вещества и радиоактивные вещества распылили, измерили, и всё такое присутствует (Indeed, there were these chemical substances and radioactive substances that were diffused, measured, and stuff like that).

(7) NUMERALS, e.g.:
Extract 15
$<$S1$>$ Результат будет всегда некрупным числом, вот и получается, допустим, две тысячи триста восемьдесят пять или тысяча шестьсот три, или что-нибудь ещё в таком роде (What we’ll have as a result will always be an odd number, like two thousand three hundred and five or one thousand six hundred and three, or something like that).

(8) QUOTATIONS, e.g.:
Extract 16
$<$S1$>$ Например, Путин недавно сказал: “Мне стыдно за бедную страну”, что-то в таком роде (For example, Putin recently said, “I’m ashamed for the poor country”, something like that).

The data indicate that in Russian professional discourse VCM exemplars are most commonly represented by clauses, while the least frequent reference is made to proper nouns (1 occurrence), verbs and numerals (2 occurrences each). The frequency of occurrence of the above exemplars is as follows (Fig. 2):
Just as the categorisation of English exemplars proved to be a challenging procedure at times, especially in terms of exemplar boundary tracing, the same impediment complicated the classification of Russian exemplars – primarily because of the word vs phrase vs clause differentiation issue. For example, in the following extract it might not at once be apparent if the exemplar is represented by a single noun проекты (projects) or by the entire clause:

Extract 17
<S1> А потом двинули обвинения, что они мол/финансировали его проекты, и все такое (And then there was this flow of accusations, alleging that they financed his projects and stuff like that).

This issue was addressed through contextual analysis, which generally provided a reasonable justification for the establishment of boundaries. Thus, in the above extract, the speaker goes on to specify other activities alongside financing, which crosses off the noun проекты (projects) from the list of potential exemplars.

As regards turn-taking, the Russian data revealed two common trends. Firstly, exemplars occurring singly tend to trigger follow-up questions more frequently than those occurring as a list. For example, in the following extract the speaker only mentions one item endangered by the new economic reform, which prompts the listener to ask a follow-up question in order to clarify/specify any other items that might be included in the list:

Extract 18
<S1> Автострахование и прочее (Vehicle insurance and all that).  
<S2> Что еще окажется под ударом? (What else is at risk?).

Secondly, corpus analysis revealed three instances of non-verbal reaction to the VCM expressed through laughter. For example:

Extract 19
<S1> С криками там “Голой капитализм!”, что-нибудь в этом духе (And yelling out things like “Down with the capitalism!”, things like that).
<S2> “Смея” (“Laughter”).
<S1> Ну это же явно нездоровое поведение, так ведь? (Well its clearly unhealthy, right?)

Thus, in Extract 19 the listener’s capacity to understand the meaning behind the utterance in general and the VCM in particular is clearly evidenced by his/her non-verbal reaction, suggesting comprehension of (and agreement with) the meaning implied by the speaker.

4.6. Functions

In terms of their functional load, Russian VCMs were revealed to perform the same basic set of functions, which makes them readily comparable with their English counterparts to create solidarity among conversational participants, or to project shared knowledge or shared identity.

A considerable difference, though, was registered in the proportion of adjunctive vs disjunctive VCMs, because the Russian analysis showed a decisive predominance of adjunctive VCMs, as opposed to the dominant set of disjunctive extenders registered in the English data. Interestingly enough, disjunctive structures are featured in the list of structural variations of Russian VCMs (see Table 4 above), and yet they are quite clearly outnumbered by their adjunctive counterparts in terms of the frequency of occurrence (78.6%). There is no obvious explanation for this difference, only to reiterate that disjunctive VCMs tend to offer more open choices for interpretation.

Another observation in terms of the functional aspect of Russian VCMs is justified by the structural analysis of the most frequently used VCM (и все такое), which also happens to be one of the shortest ones in the list. This preferred brevity of expression, as we believe, is indicative of yet another function of VCMs – a time-saving one – which conforms to an overall principle of linguistic economy.

5. Discussion and conclusions

5.1. The syntax of vague category marking

By looking at VCMs in both English and Russian, it is clear that, while there may be formal features which are difficult of direct translation (especially, perhaps, the minor variations which any core VCM phrase can display), both languages appear to have VCMs as part of their repertoire of chunks or fixed expressions. This is evidenced in the fact that in both languages a VCM may be attached to a variety of syntactic types of exemplar, thus indicating the absence of an internal grammar.

VCMs in both languages are predominantly noun- and pronoun-based (with the exception of etcetera, and so on/so forth and their translation equivalents), yet there is a strong tendency for exemplars to include verb phrases and clauses. This is
perhaps not surprising, given the nature of business and professional communication, where states, activities and processes are always to the fore, alongside references to nominal entities. These varied syntactic contexts of marking are facilitated by the polysemy and weak collocability of nouns such as thing and stuff. Quirk et al. (1985) observe that thing, although a noun, functions like a pro-form, “conveying a broad and undifferentiated meaning” (p.77), which is the case with the Russian language as well.

In both languages, it was found that precisely establishing the boundaries of exemplars was on occasion problematic, though there is no evidence that understanding what constituted an exemplar was ever problematic for the listener(s). That there may be more than one exemplar surrounding the VCM (e.g. extracts 6 and 7) means that an exemplar is not necessarily to be interpreted as the prototype of its class. Table 3 showed that, for English, verb phrases and clauses acting as exemplars (the most indeterminate category) were one of the two most frequent exemplar types. The same applied in Russian, where clauses constitute the single most common type of exemplar to proceed the selected VCMs.

5.2. Functional and pragmatic aspects

The functions of the VCMs analysed from the Russian data largely correspond to those present in the English data, that is to say, the projection of non-lexicalised, fluid and instantial categories which both protect the speaker from over-explicit commitment and which tactically bond with the listener to create and maintain shared membership of individual enterprises, to create and consolidate business relationships and to project membership of the broad professional community. They are important markers of identity: one has to be a member of the appropriate group (whether commercial company or professional grouping) fully to apprehend many of the instantaneous, fluid categories being invoked at the time. Here the evidence of turn-taking and the listener responses are crucial. Extracts 3, 4, 9, 18 and 19 all show how the listeners’ reactions evidence both the comprehension and acceptance of the VCM but also underline the nature of vague category marking as a co-constructed enterprise: the VCM does not hang in vacuo at a turn-ending or as a mere coda but is an invitation to consent to its function at that moment in the discourse.

Clearly important in the field of business negotiation or professional discussion are the strategies of hedging and face-protection. Extracts 4, 5 and 9 for English and 11, 14 and 15 for Russian may be seen as protecting the speaker against the face threats immanent in over-specification.

In any context of negotiability, whether it be in the exploration of ideas or processes or the negotiation of business deals and relationships, it would seem that VCMs have developed a pragmatic specialization, with adjunctive VCMs projecting the instantial categories as mutually apprehended, while disjunctive VCMs leave more doors open and the categories remain expandable or contractable. The Russian analysis revealed a preference for adjunctive VCMs while the English data had more disjunctive types. This is probably due to the more strictly circumscribed business contexts of the English data as opposed to the wider professional contexts included in the Russian data, though it may also reflect cultural norms, an assertion untestable without more closely matched corpora. Whatever the case, pragmatic specialization of VCMs enables users to project and consolidate particular types of shared identity within groups of interlocutors.

5.3. Data issues

This article has raised and sought to address problems associated with the comparability of corpus data across languages, both in terms of the selection and translatability of search terms as well as technical issues of corpus size, text-type and data annotation: for example, annotated pauses can support the interpretation as to where a VCM event ends, while the indication of latched utterances in the middle of a VCM event may support an interpretation of convergence on the part of the listener(s). The ability to carry out cluster analyses is also a useful facility not always shared by corpora and their associated search engines. The two datasets used here do not constitute a parallel corpus in any sense but do offer similar settings from comparable communities of practice (Wenger, 1998) and thus provide a resource for an initial, tentative investigation which may, someday we would hope, be enhanced by more closely matched data sources.

Acknowledgements

The publication was prepared with the support of the “RUDN University Programme 5-100”.

References


[---]