Chapter 13
Applying Corpus Linguistics to a diagnostic tool for pain

Elena Semino, Andrew Hardie, Lancaster University, UK and Joanna Zakrzewska, Eastman Dental Hospital and UCLH NHS Foundation Trust, UK

1. Introduction

Pain is notoriously difficult to diagnose and treat, especially when it has become chronic. Numerical scales are often used to assess the *intensity* of pain, i.e. how severe it is (e.g. on a scale from 1 to 10), while linguistic descriptors are included in some diagnostic tools (e.g. the LANSS Pain Scale, Bennett, 2001) that aim to capture the *quality* of the pain, or what it feels like (e.g. 'hot' vs 'cold'). In this chapter, we focus on one particular tool for the diagnosis of pain, the McGill Pain Questionnaire, which exploits 20 groups of linguistic descriptors to capture both the quality and intensity of pain experiences (e.g. 'hot', 'burning', 'scalding', 'searing', in assumed order of increasing intensity) (Melzack, 1975).

The wide use of the McGill Pain Questionnaire (hereafter MPQ) for the diagnosis of pain is supported by evidence of its general validity and reliability for a range of conditions. However, as we explain below, a number of studies have identified issues that cast doubt on the appropriateness of the particular set of linguistic descriptors included in the MPQ, and on the ways in which they are grouped in the questionnaire's internal structure. We suggest that these problems are due to the way in which the descriptors were originally selected, and show how the methods of corpus linguistics – the computer-aided study of language on a large scale (McEnery and Hardie, 2012; see also Brookes, this volume) – can be used to investigate systematically the potential linguistic problems within the MPQ. We also provide some evidence of how these linguistic problems can affect patients' responses in ways that undermine the validity of some aspects of the questionnaire.

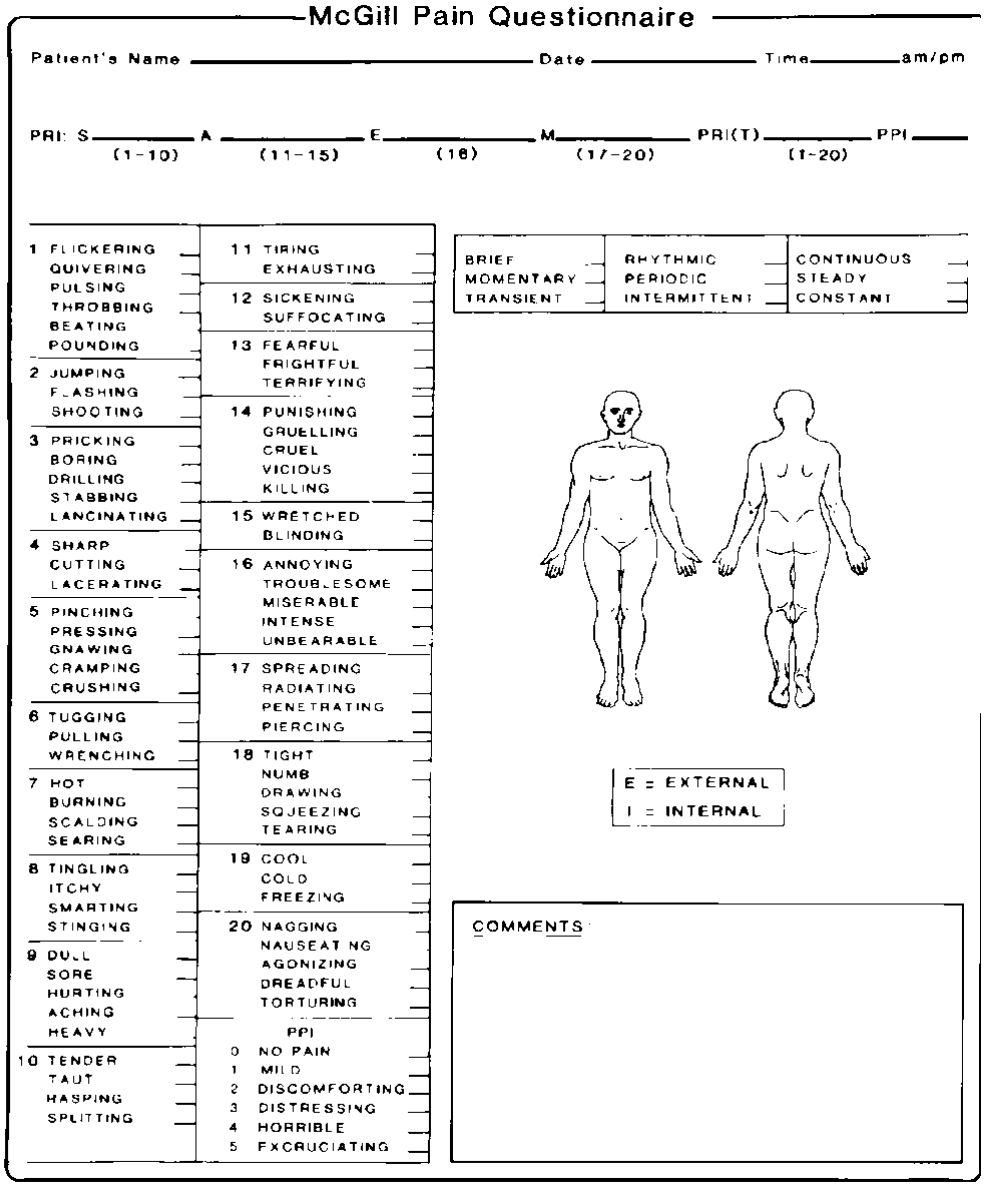
In section 2 we introduce the MPQ and provide an overview of the clinical studies that have considered the problems that may be associated with the linguistic descriptors it includes. In section 3 we briefly outline our method, before applying some central methods in corpus linguistics to investigate the possible causes of some of the issues identified in the literature in section 4, and more generally to provide a systematic account of potentially problematic aspects of variation among the descriptors in the MPQ. In section 5 we test the influence on patients' selections of one particular aspect of variation: the extent to which each descriptor tends to be used to describe pain in English generally. By analysing the selections made by 800 patients, we show how variation in the strength of the association between each descriptor and the word 'pain' undermines the validity of the MPQ as a way of measuring pain intensity, at least in some parts of the questionnaire. In section 6 we draw some conclusions on the implications of our findings for the use of the MPQ, and for the assessment and design of language-based diagnostic tools for pain more generally, before providing some concluding remarks. We propose that such tools need to be based on

systematic linguistic evidence, and suggest that corpus linguistic techniques are ideally suited to provide such evidence.

2. The McGill Pain Questionnaire

The MPQ (reproduced as Figure 1) includes, among other things, 78 one-word pain descriptors in English, divided into 20 groups (Melzack, 1975). The groups capture different qualities of pain experience and fall into four broader classes, depending on the aspect of pain they relate to: sensory (groups 1-10), affective (groups 11-15), evaluative (group 16) and miscellaneous (groups 17 to 20). Each group contains between 2 and 7 descriptors, listed in order of increasing intensity of pain. For example, group 4 captures the sensory quality of Incisive Pressure and contains three descriptors: 'sharp', 'cutting' and 'lacerating'. 'Sharp' is the descriptor associated with lowest intensity within the group, and lacerating with the highest intensity.[i] Patients therefore have the option of choosing either a single descriptor from each group, or else none, if that particular quality of pain does not apply to them. For example, if a patient's pain does not feel hot, they do not select any descriptors from group 7.

Figure 1 – The McGill Pain Questionnaire (Melzack, 1983: 44)



In the seminal paper that first described the MPQ, Melzack (1975) explains that the linguistic descriptors were drawn from his own experience of the language used by patients and from the literature on pain, including previously compiled lists of pain descriptors. Physicians and other university graduates were involved in dividing descriptors into groups, while the intensity scales were determined on the basis of scores provided by groups of doctors, patients and students. Melzack (1975) also reported some initial evidence for the validity and reliability of the MPQ, and for its ability to detect differences between different methods of pain relief.

Since its introduction, the MPQ has become the most widely used diagnostic tool for pain. It has been translated into 26 different languages (Main, 2016) and has been found to be a valid, reliable and sensitive tool in clinical settings involving patients with a variety of

different conditions, from low-back pain (Melzack et al., 1985) to cancer (Ngamkham et al., 2012). As Main (2016: 1390) puts it, 'there is evidence that the MPQ 1) can discriminate between pain conditions, and also capture variation within conditions, 2) is sensitive to change, and 3) is responsive to treatment and can be used as an outcome measure'.

However, a number of studies have also identified weaknesses that relate specifically to the choice of descriptors in the questionnaire. More specifically:

- Some descriptors have been found to be difficult for patients to understand (e.g. 'rasping' and 'lancinating');
- Some descriptors have multiple meanings, and may therefore be ambiguous (e.g. 'boring', which is included in the Punctate Pressure sensory group);
- Some descriptors do not clearly belong in the group within which they have been placed (e.g. 'heavy', which is included in the Dullness group);
- Some descriptors are not clearly relevant to pain, and are seldom selected by patients (e.g. 'rasping'). (see Fernandez and Towery, 1996; Wilkie et al., 2001; Fernandez and Boyle, 2002; and Main, 2016 for a systematic review).

Main (2016: 1394) suggests that, as a result of these issues, patients' choice of descriptors may be influenced by general language use, their verbal repertoire and level of health literacy, and what he refers to as 'a vulnerability to "effort after meaning".' The way in which the questionnaire is administered may also affect patients' choices. Some clinicians read out each group to patients, giving them little time to ponder over their choices, while, in other cases, patients fill in a written version of the questionnaire in their own time.

The problems with the set of descriptors included in the MPQ have so far been addressed in different ways. Two studies have suggested a reduction in the overall number of descriptors on the basis of Psychology students' reclassification of the existing set into groups (Fernandez and Towery, 1996; Fernandez and Boyle, 2002). Lin et al. (2011) identified the most useful descriptors in the MPQ by comparing the full set with the words used by patients with spinal cord injury and multiple sclerosis when asked to describe their pain in telephone interviews.

Finally, two shorter versions of the MPQ have been produced: Short-form-MPQ (SF-MPQ; Melzack, 1987) and Short-form-MPQ-2 (SF-MPQ-2; Dworkin et al., 2009). SF-MPQ was developed as a less time-consuming version of the original MPQ, and contains 15 descriptors, each rated on a 4-point intensity scale. SF-MPQ-2 contains 7 additional items intended to be relevant to neuropathic pain (i.e. pain caused by problems in the nervous system), and adopts a 10-point scale for pain intensity.

While patients' language use was taken into account in some way in the development of the MPQ and both short forms, none of the versions was based on systematic large-scale linguistic evidence. In the next sections, we show how corpus linguistic methods can be applied to gain a better understanding of the problems with the original MPQ that have been pointed out in the clinical literature.

3. Methodology: corpus linguistics

Corpus linguistics involves the use of specialised software to study large collections of texts known as 'corpora' – the plural of 'corpus' (McEnery and Hardie, 2012; see also Brookes,

this volume, for an overview of corpus linguistics). By 'large' we mean collections of texts that could not be read and analysed by hand and eye alone within reasonable time scales. In practice, this can range from a few hundred thousand words for specialised corpora that only include texts from a particular genre (e.g. a corpus of UK party manifestos from the 20th century), to billions of words for corpora that aim to represent a whole language. The MPQ data, of course, is itself unsuitable for corpus-based analysis: it is far too short, and the responses consist of isolated descriptors rather than natural running text. However, exploring how the descriptors are used outside of the MPQ, in large corpora of general English, can provide evidence that is relevant to an assessment of potential weaknesses in the questionnaire.

In this chapter, we make use of the Oxford English Corpus (hereafter OEC), which includes 2.5 billion words of 21st century English. It is mostly based on material collected from the World Wide Web, but includes texts from a wide variety of genres and 'domains': e.g. news and media, law, medicine, science, business, fiction, personal blogs. It also contains texts from international varieties of English from all parts of the world (e.g. UK, US, Australia, India, Singapore), and is therefore a good reference corpus of 'general English'. We used the Sketch Engine software (Kilgarriff et al., 2014; https://www.sketchengine.eu) to study the use of the 78 descriptors in the OEC, particularly in terms of their frequencies and the extent to which they tend to be used to describe pain. We describe the specific corpus linguistic techniques we used in more detail in the relevant sections.

4. Corpus linguistic methods and the pain descriptors in the MPQ

The structure of the MPQ relies on two dimensions of variation among the 78 linguistic descriptors of pain. *Across* groups, the descriptors are intended to vary according to the *quality* of the pain experience, whether in sensory, emotional or evaluative terms. *Within* each group, the descriptors are intended to vary according to the *intensity* or severity of that particular quality of the pain experience. The issues identified in the clinical literature suggest that there may be problems in how the descriptors were selected and allocated to groups. From a linguist's perspective, it is in fact not at all surprising that any set of 78 words will involve differences across many more dimensions than the two that are assumed in the questionnaire. In this section, we use corpus linguistic methods to provide evidence of some of those dimensions of variation. In the next section, we take one of those dimensions of variation and show to what extent and how it can interfere with the intensity scales when patients fill in the MPQ.

4.1 Frequency in English

A first step was to consider the overall frequency of the descriptor words in English. The corpus frequency of a word can be considered a proxy measure for its familiarity to a hypothetical average speaker. The question, then, is whether the corpus data suggests that the descriptors are likely to be equally or unequally familiar to that hypothetical average speaker. We began by obtaining the raw number of occurrences of the 78 MPQ descriptors in the OEC as a whole – that is, counting all instances of each descriptor, whether they are used in the context of pain or not . The complete set of figures is provided in Table 1, where

the descriptors are sorted in decreasing order of frequency. The table also provides the relative frequencies of each descriptor in the corpus per million words.

Table 1 – Frequencies of MPQ descriptors in the OEC, in descending order.

Group	Descriptor	OEC Raw frequencies	OEC Normalized frequencies (per million words; three decimal places)
7	Hot	206291	84.857
19	Cold	190057	78.180
9	Heavy	169869	69.875
19	Cool	158752	65.302
14	Killing	136171	56.014
18	Drawing	112151	46.133
2	Shooting	104238	42.878
4	Sharp	88072	36.228
4	Cutting	84525	34.769
18	Tight	78879	32.447
16	Intense	67234	27.657
7	Burning	56137	23.092
1	Beating	52764	21.704
6	Pulling	49155	20.220
10	Tender	40452	16.640
3	Boring	39278	16.157
5	Pressing	32809	13.496
17	Spreading	31437	12.931
9	Dull	29701	12.217
14	Cruel	29292	12.049
2	Jumping	28662	11.790
16	Annoying	25821	10.621
16	Miserable	21137	8.695
14	Vicious	21106	8.682
9	Sore	19005	7.818
19	Freezing	17757	7.304
13	Terrifying	14750	6.067
20	Dreadful	14610	6.010
3	Drilling	14442	5.941
9	Hurting	14345	5.901
13	Fearful	13268	5.458
2	Flashing	11983	4.929
5	Crushing	10924	4.493
10	Splitting	10737	4.417
1	Pounding	10555	4.342

18	Tearing	10436	4.293
17	Piercing	8353	3.436
14	Punishing	7657	3.150
16	Troublesome	7362	3.028
3	Stabbing	7344	3.0209
18	Numb	6945	2.857
15	Wretched	6866	2.824
16	Unbearable	6770	2.785
17	Penetrating	6393	2.630
18	Squeezing	6124	2.5191
11	Exhausting	5915	2.433
20	Nagging	5782	2.379
9	Aching	5093	2.095
15	Blinding	4782	1.967
11	Tiring	4752	1.955
8	Stinging	4624	1.902
20	Agonizing	4460	1.835
12	Sickening	4372	1.798
14	Gruelling	4201	1.728
7	Searing	4005	1.647
1	Flickering	3752	1.543
20	Torturing	3734	1.536
1	Throbbing	3647	1.500
10	Taut	3378	1.389
8	Itchy	2759	1.135
6	Wrenching	2610	1.074
17	Radiating	2511	1.033
12	Suffocating	2497	1.027
6	Tugging	2282	0.939
1	Quivering	2169	0.892
8	Tingling	2105	0.866
1	Pulsing	1923	0.791
5	Pinching	1887	0.776
20	Nauseating	1713	0.705
5	Gnawing	1615	0.665
13	Frightful	1568	0.644
8	Smarting	1223	0.503
5	Cramping	1049	0.431
7	Scalding	962	0.396
10	Rasping	961	0.395
3	Pricking	573	0.236
4	Lacerating	312	0.128
3	Lancinating	15	0.006

As Table 1 shows, the frequencies of the 78 descriptors in English span across a very wide range, between 206,291 occurrences for 'hot' and 15 occurrences for 'lancinating'. At the top of Table 1, we find some very common English adjectives alongside 'hot', such as 'cold', 'heavy' and 'cool'. The 22 most frequent descriptors in the MPQ occur more than 10 times per million words in the OEC, and seven of them are frequent enough to have been included in the New General Service List (New-GLS) of English vocabulary. The New-GLS is a list of 2,494 words that make up over 80% of the word tokens in four large corpora of contemporary English, and that can therefore be used as a basis for teaching materials for non-native speakers of English (Brezina and Gablasova, 2015). The seven MPQ descriptors that are included in that list are: 'hot', 'heavy', 'cold', 'sharp', 'cool', 'tight', and 'intense'. In 15 further cases, the New-GLS includes a word that is derivationally and semantically related to an MPQ descriptor, often being part of what linguists call the same 'lemma' (e.g. 'cut' and 'cutting' can be described as being part of the same lemma, and would be subsumed under the same entry in a dictionary). In Table 2, these 15 cases are listed alongside the relevant word in the New-GLS, in decreasing New-GLS rank order.

Table 2 – MPQ descriptors and closely related words in the New-GLS

MPQ descriptor	**Related word and word class in New-GLS**
cutting	cut (verb)
pulling	pull (verb)
killing	kill (verb)
fearful	fear (noun)
troublesome	trouble (noun)
pounding	pound (noun)
pressing	press (verb)
beating	beat (verb)
shooting	shoot (verb)
splitting	split (verb)
burning	burn (verb)
jumping	jump (verb)
hurting	hurt (verb)
sickening	sick (adjective)
tearing	tear (verb)

At the bottom of Table 1, by contrast, we find words that can be described as rare in English. The 15 descriptors at the bottom of Table 1 (e.g. 'quivering', 'smarting' and 'taut') occur less than once per million words in the OEC. In addition, some of these descriptors are not spread evenly in the corpus, but tend to occur in particular, often specialised, genres. About a third of the instances of 'taut', for example, occur in art criticism, in uses such as 'the bandstand's giant winged canopy echoes the taut exuberance of the Pavilion', from an architectural review. A particularly extreme case of specialisation in use is 'lancinating'. With 15 occurrences, it is the least frequent MPQ descriptor in the OEC. Moreover, all 15 instances occur in scientific papers that discuss the MPQ. In other words, there is no

evidence that ‘lancinating’ has any currency in present-day English outside of the MPQ itself.

Overall, therefore, the 78 descriptors in the MPQ vary considerably in terms of frequency in English, as represented in the OEC, and therefore in the extent to which they are likely to be familiar words to patients who fill in the questionnaire. More specifically, a subset of the descriptors are rare words, sometimes associated with specialised genres, and may therefore be unfamiliar to many patients, depending on the breadth of their verbal repertoire. This begins to explain the observation in the clinical literature on the MPQ that some descriptors are difficult to understand, with ‘lancinating’ being the most extreme example of this difficulty (see Fernandez and Towery, 1996).

While in this section we have discussed the general frequencies of the 78 descriptors in the OEC, in the next section we consider more specifically the extent to which they are used to describe pain in the corpus.

4.2 Prevalence of pain-related uses

The fact that a particular descriptor may be a frequent and familiar word in English does not necessarily mean, however, that it is also a good candidate for inclusion in a pain questionnaire such as the MPQ. Indeed, many of the adjectives at the top of Table 1 capture important properties of objects (e.g. temperature and weight), but, even at an intuitive level, are unlikely to be perceived as strongly associated with pain (e.g. ‘heavy’). We therefore used the OEC to investigate the strength of the association between each descriptor and the experience of pain. Our approach was twofold. First, we analysed the 50 top ‘collocates’ of each descriptor in the OEC (see below for more detail). Second, we computed the number of times that the word ‘pain’ occurs in close proximity to each descriptor in the corpus. Table 3 in the Appendix summarises the results of the analysis. In the table, the 78 descriptors are sorted in order of decreasing frequency of how often the word ‘pain’ occurs in a window 5 words to the left and 5 words to the right of an occurrence of that descriptor in the OEC. Consider, for example, this short extract from the OEC: ‘who can forget that sharp, stinging pain followed by the itchy swelling’. Here ‘pain’ occurs two words to the right of ‘sharp’. We now explain each analysis in more detail.

4.2.1 Top collocates of each MPQ descriptor in the OEC

The notion of ‘collocation’ in corpus linguistics captures the tendency of words to occur together with other words (see Brezina et al., 2015). This tendency is important because it is likely to have psychological consequences. For example, after reading a word such as ‘fish’, people tend to be faster at recognising a word such as ‘chips’ than words such as ‘car’ or ‘hammer’ (Hoey, 2005; Hughes and Hardie, in press). This is because ‘fish and chips’ is a frequent and familiar phrase in (British) English, so that the strength of the association between the two words in language use corresponds to a strong association between the two words in people’s minds.

Within corpus linguistics, collocational relationships are described quantitatively according to a number of different factors. The extraction and analysis of collocations in corpus data is often based on or takes account of the following four aspects:

- Distance: the span around the word of interest (or 'node' word) in which other words (the 'collocates') occur, such as 5 words to the left and 5 words to the right.
- Frequency: how often a particular collocate is used in close proximity with the node word in a particular corpus; the frequency of the node/collocate combination can be interpreted both as a measure of how much evidence there is for the reality of that collocation, and of how likely speakers of the language are to be familiar with the co-occurrence pattern in question.
- Exclusivity: the likelihood that, given the node word, the collocate will occur in close proximity to it; when this is high, the node and collocate can be considered as 'exclusive' to one another; when it is low, both also occur with many other words as well.
- Distribution: whether the combination of the node word and a given collocate is restricted to a small number of texts, or to a particular genre, as opposed to being found widely distributed across many different texts and genres.

For example, Brezina et al. (2015) compare the words 'in' and 'affair' as collocates of the word 'love'. The combination of 'in' and 'love' is frequent but not exclusive: the phrase 'in love' is common in English (45,795 instances in the OEC), but 'in' occurs in close proximity to many other nouns in addition to 'love'. In contrast, the combination of 'affair' and 'love' is less frequent (5,160 instances in the OEC) but more exclusive: if 'affair' is used, there is a high probability that 'love' will precede it. Considering again 'lancinating', the relationship between this descriptor and the word 'pain' is strong in terms of exclusivity and weak in terms of frequency: if 'lancinating' occurs, there is a very high probability that 'pain' will occur in close proximity, but these co-occurrences are extremely infrequent. Moreover, these co-occurrences are very narrowly distributed – not only do all examples occur in a very specialised genre in the OEC, they also all occur in texts on a specific topic (that of the MPQ itself).

We used the Sketch Engine software to compute the collocates of each of the 78 MPQ descriptors in the OEC. We opted for a span of 5 words to the left and 5 words to the right of each node word, and ranked the collocates using the Dice coefficient (which in the Sketch Engine is implemented as 'logDice' – see Rychlý, 2008). While dozens of statistical measures for scoring collocations have been proposed, we selected the Dice coefficient because it primarily reflects the strength of association between node and collocate (that is, exclusivity) but, unlike some other measures of association strength, does not unduly emphasise very infrequent combinations. This, in effect, means that this part of our analysis balances the frequency and exclusivity criteria. We considered the top 50 collocates of each descriptor, and provide the top ten collocates in Table 3 in the Appendix.[ii]

Nine of the descriptors have 'pain(s)' as one of the top 50 collocates: 'dull', 'nagging', 'sharp', 'intense', 'lacerating', 'unbearable', 'lancinating', 'stabbing' and 'agonizing'. In all cases, in the collocation the descriptor is used to describe the pain itself (e.g. 'agonizing pain'). 'Exhausting', in contrast, has 'painful' as one of its top collocates, but here the collocation seems to be due to the fact that the two words are used to describe different aspects of the same experience, as in 'IBS is a painful and exhausting disease'.

Seven of the descriptors have 'ache(s)' or 'aching' as one of the top 50 collocates, including three from the previous list: 'numb', 'dull', 'throbbing', 'sore', 'nagging', 'unbearable', and 'gnawing'. These adjectives tend to function in context as pain descriptors, or, in some instances of 'numb', as descriptors of unpleasant experiences that

occur together with pain, as in 'my aching legs and numb feet'. In total, therefore, *only 13* of the 78 descriptors in the MPQ have among their top 50 collocates a word that describes pain generally.

In a number of other cases, the top 50 collocates include words that describe or are associated with pain in a particular part of the body. Seven descriptors have 'headache(s)' as one of the top 50 collocates, including some that we have mentioned in the previous paragraph: 'throbbing', 'sore', 'splitting', 'pounding', 'dreadful', 'nagging', 'aching'. Of these, 'splitting' and 'pounding' have 'headache(s)' as one of their top 10 collocates, as shown in Table 3 in the appendix. This suggests such a strong association with headaches that we might well speculate that these two descriptors are not appropriate for other kinds or sites of pain. The top 50 collocates of seven further descriptors include words that refer to physical damage or to illness and symptoms of illness, as shown in Table 4.

Table 4 – Seven MPQ descriptors and top 50 collocates that refer to physical damage and illness.

MPQ descriptor	**Relevant collocates**
cramping	bloating, diarrhoea, nausea, constipation, colicky
nagging	injury, injuries
itchy	scabies
tingling	neuropathy and paresthesia
radiating	sciatica
agonizing	spasm(s) and asphyxiation
blinding	trachoma
smarting	wounds

Finally, 18 descriptors have body parts among the top 50 collocates, as shown in Table 5.

Table 5 – 18 MPQ descriptors and top 50 collocates that refer to body parts.

MPQ descriptor	**Relevant collocates**
troublesome	groin, hamstring
quivering	lip(s), nostrils
pulsing	veins
pinching	nipple(s)
taut	muscles, abs, tummies
throbbing	cock, temples
wrenching	gut
sore	throat(s), thumb, muscles
splitting	hairs
pounding	fist, chest
hurting	knee, ankle
nagging	hamstring, ankle

itchy	scalp
tingling	tastebuds
lacerating	tongue
gnawing	incisors, fingernails
piercing	nipple, genital, armour, ear, navel
aching	joints, muscles, limbs, legs

Some of the collocations in Tables 4 and 5 indicate specialisation in the type or site of pain, as in 'tingling' and 'neuropathy' or 'aching' and 'joints'. Some describe unpleasant sensations but not pain specifically, as in 'itchy' and 'scabies'. Some are to do with physical sensations in very specific contexts where pain may be relevant but is not necessarily central, e.g. sex in the case of 'pinching' and 'nipples'. And some are not to do with pain at all. This applies to 'taut' and 'muscles', for example, and to the pairs of collocates that tend to co-occur in figurative expressions, as in the case of 'lacerating' and 'tongue'.

Finally, the top 10 collocates of the 78 descriptors (provided in Table 3 in the appendix) suggest that a number of descriptors in the MPQ have much stronger associations with non-pain-related words than with any words that relate to pain. For example, the top 10 collocates of 'drawing' suggest that it is mostly used in its art-related meaning (cf. 'paintings', 'painting', 'board', 'prints', 'photographs', 'pencil', and 'ink'). Similarly, the top 10 collocates of 'boring' show that it is mostly used in the sense of 'uninteresting' (cf. 'repetitive', 'bland', 'predictable', 'tedious', 'mind-numbing', and 'plain').[iii] In cases such as 'boring', this can potentially lead to confusion, as the more frequent meaning could interfere with the pressure-related meaning that is relevant to the MPQ when patients fill in the questionnaire.

Overall, therefore, our analysis of the top collocates of the 78 MPQ descriptor shows a great deal of variation in the extent and nature of the association between each descriptor and the verbal expression of pain experiences.

4.2.2 Co-occurrences with *pain* in the OEC

Our second approach to collocation involved calculating the raw number of occurrences of the word 'pain' (in the singular or plural) within a span of 5 words to the left and 5 words to the right of each instance of one of the 78 MPQ descriptors in the OEC. This was a way of investigating to what extent each of these words is used as a potential pain descriptor in the corpus, regardless of whether any words to do with pain occur among its top collocates according to logDice or any other potentially relevant measure of statistical significance. In Table 3 (see Appendix), the descriptors are listed in descending order of number of co-occurrences with 'pain' within that span.

As with the overall frequencies of occurrence of each descriptor in the OEC, the range of co-occurrences with 'pain' shown in Table 3 is very wide. At the top, 'sharp' has 986 co-occurrences with 'pain', while 'rasping', at the bottom, has none. More specifically, 24 descriptors have 10 or fewer instances of 'pain(s)' within the relevant span in the whole corpus: 'flashing', 'jumping', 'sickening', 'quivering', 'pricking', 'lacerating', 'tugging', 'penetrating', 'nauseating', 'scalding', 'taut', 'drilling', 'pinching', 'frightful', 'exhausting', 'torturing', 'smarting', 'suffocating', 'gruelling', 'flickering', 'tiring' and 'rasping'. An examination of the context in which pain co-occurs with these and other descriptors also shows that, in a number of cases, the descriptor is not used to describe what the pain feels

like, but rather the cause of the pain (as in the case of 'beating', 'drilling' and 'tearing') or an unpleasant sensation that is experienced at the same time as the pain, as we already noted with 'exhausting' above (as in the case of 'tingling' and 'sickening').

Overall, therefore, a corpus-based examination of the use of the 78 MPQ descriptors in a large corpus of present-day English shows that they vary along more potentially relevant dimensions than is assumed, and in fact required, by the structure of the MPQ. In particular, some (and possibly many) of the descriptors are infrequent enough to be unfamiliar to at least some patients. Some have a very weak association with pain and a few almost never occur in a context where pain is talked about. Of those that do occur in pain-related contexts, some do not describe the pain as such, while others very strongly specialise in the description of pain in a particular part of the body or resulting from a particular problem. Only a relatively small number are shown by the corpus analysis to be frequent and general pain descriptors.

These findings can begin to explain some of the problems discussed in the clinical literature, such as lack of comprehension or confusion in patients' reactions to some of the descriptors. To provide further evidence of the relevance of our corpus-based observations, in the next section we report on a pilot study that was conducted to test whether the strength of the collocation between of each descriptor and 'pain' in the OEC could, in fact, affect patients' responses in a way that undermines the aims of the MPQ.

5. Strength of collocation and the intensity scales in the MPQ: a pilot study

As we mentioned above, the strength of the collocation between two words in language use has been related, both theoretically and empirically, to the psychological association between the two words. In particular, these 'priming effects' have been observed between pairs of collocates, so that exposure to one member of the pair leads to faster recognition of the other member of the pair in experimental settings. Like the phenomenon of collocation itself, this is often considered in terms of activation spreading: units in the mental lexicon are connected in a network, such that when one unit is used (comprehended/produced) its mental representation is activated, and that activation partially spreads to the other units most directly connected to it – making them more readily available for subsequent comprehension or production themselves.

It can therefore be hypothesised that the variation in collocation strength between each MPQ descriptor and 'pain', or semantically similar words, could produce a biasing effect when patients fill in the questionnaire. As the mental representation of pain will be, we can assume, strongly activated in this situation, the collocates of 'pain' will themselves be pre-activated, and thus more accessible, i.e. likely to be produced, and easier and quicker to comprehend. This is particularly problematic for the intensity scales within each group, where patients' selection of one out of several descriptors is often interpreted as reflecting only the level of severity of the pain the descriptor refers to. We therefore set out to test the potential interference of collocation strength with intensity ratings in the MPQ.

We used the figures for the number of occurrences of 'pain' in a +5/-5 span around each descriptor in the OEC (in Table 3 in the Appendix and discussed above) to calculate the probability that a given instance of 'pain' co-occurring with any one of the 78 descriptors involves a particular descriptor. So, for 'hot', for example, we determine the probability that

– for any particular OEC instance of 'pain' that we know co-occurs with *some* descriptor – the descriptor in question is 'hot' rather than one of the other 77 possibilities. Since there are 8,691 total instances of 'pain' in the OEC within this span of any descriptor, this probability is equal to the raw number of co-occurrences for a particular descriptor divided by 8,691. In corpus linguistic terms, this is an asymmetric measure of the strength of the collocation: it measures the strength of the link from 'pain' to the descriptor, and ignores the strength of the link from the descriptor to 'pain'. This is in contrast to the Dice coefficient, used above, which is a symmetric measure (effectively averaging the strengths of the two directions of association). We then examined patients' selections in 800 questionnaires that were filled in at the Eastman Dental Hospital in London, UK. All patients presented with orofacial pain that lasted more than three months without an identifiable cause. All patients completed the MPQ in a consultation with the same doctor. In all cases, the doctor read out each group of descriptors one by one, and paused after each group to allow the patient to select one or no descriptor from that group, as appropriate to their pain.

We considered each group of descriptors separately and looked at the level of correlation between the probability that each descriptor is chosen on the one hand and, on the other hand, the strength of the collocation between each descriptor and the noun *pain*, as described above. Pearson correlation coefficients were calculated for all groups of descriptors in the MPQ. Here we will focus specifically on the sensory groups that have a Pearson correlation coefficient above 0.9, i.e. groups 1, 2, 3, 4, 6, 7 and 8 (see Table 5). Generally in analysis of correlation, a Pearson co-efficient of 0.7 or higher is interpreted as indicating a strong relationship; >0.9 thus represents a *very* strong relationship (one in which >81% of the variability in one variable is explained by the value of the other variable).

Table 5 – Pearson coefficients for correlation between the probability that each descriptor is chosen and the strength of the collocation between the descriptors in each MPQ group and the word 'pain': sensory groups with coefficients >0.9.

MPQ group	**Descriptors**	**Pearson correlation coefficient (3 decimal places)**
1	flickering, quivering, pulsing, throbbing, beating, pounding	0.960
2	jumping, flashing, shooting	0.998
3	pricking, boring, drilling, stabbing, lancinating	0.991
4	sharp, cutting, lacerating	0.997
6	tugging, pulling, wrenching	1.000
7	hot, burning, scalding, searing	0.991
8	tingling, itchy, smarting, stinging	0.985

As shown in Figures 2-5 below, groups 1, 2, 3 and 4 share a similar pattern. All four groups include one very strong collocate of 'pain', in contrast with all the other descriptors in the

group, and patients overwhelmingly select the strong collocate, i.e. ‘throbbing’ in group 1, ‘shooting’ in group 2, ‘stabbing’ in group 3, and ‘sharp’ in group 4. This is regardless of where the strong collocate falls in the intensity scale in the relevant group: ‘throbbing’ is in the middle, ‘shooting’ and ‘stabbing’ are at or near the top, and ‘sharp’ is at the bottom.

Figure 2 – Frequency of patient selections and frequency of co-occurrence with ‘pain’ in the OEC for Group 1 descriptors in the MPQ.

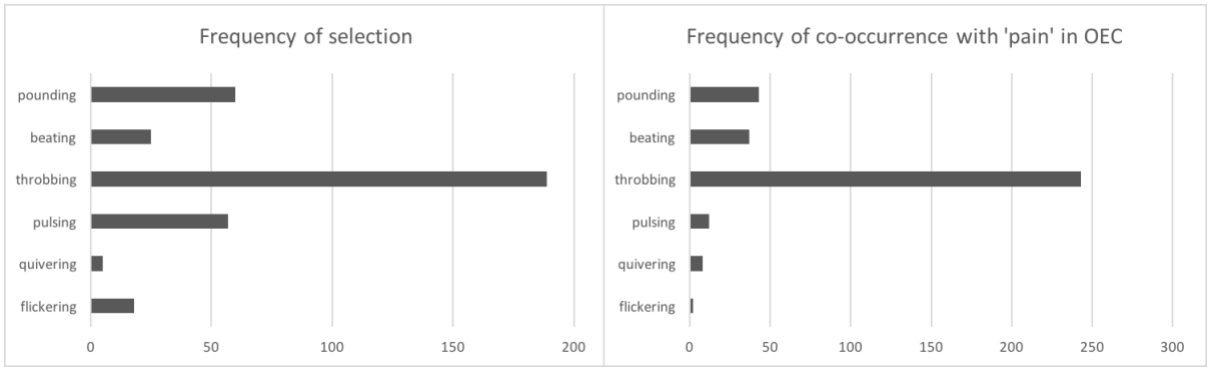


Figure 3 – Frequency of patient selections and frequency of co-occurrence with ‘pain’ in the OEC for Group 2 descriptors in the MPQ.

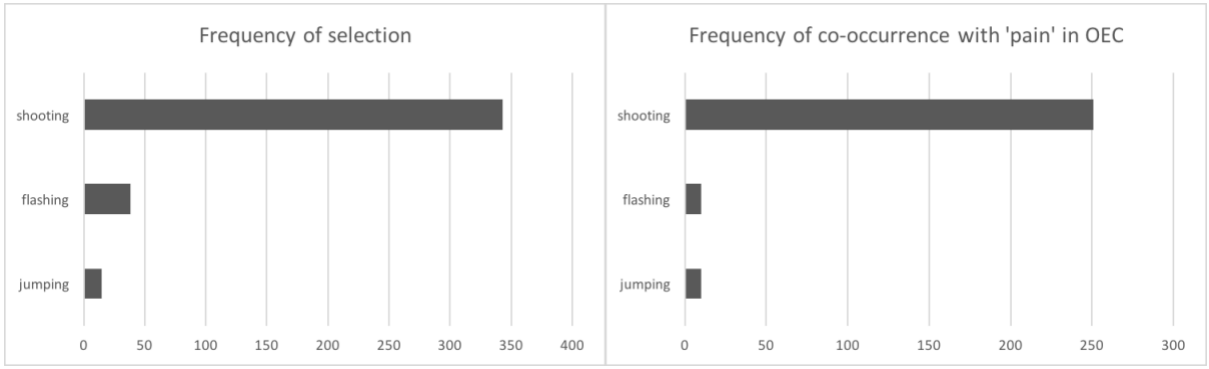


Figure 4 – Frequency of patient selections and frequency of co-occurrence with ‘pain’ in the OEC for Group 3 descriptors in the MPQ.

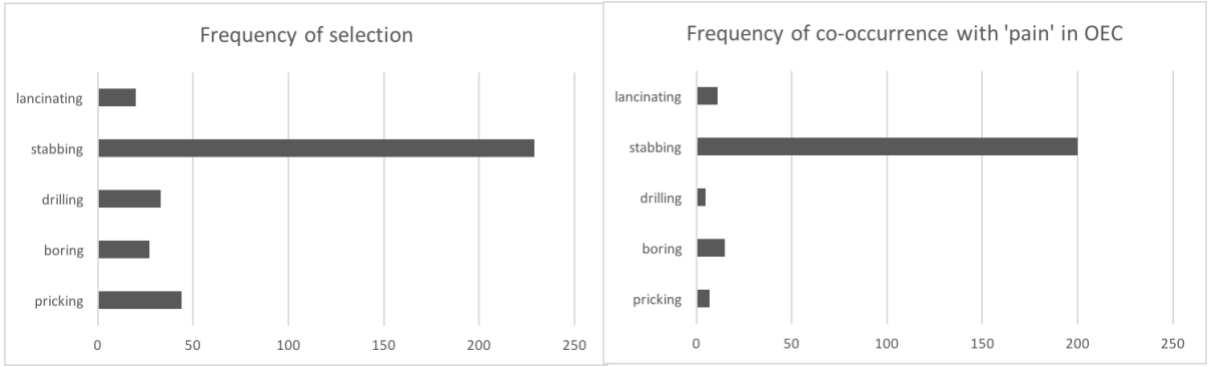
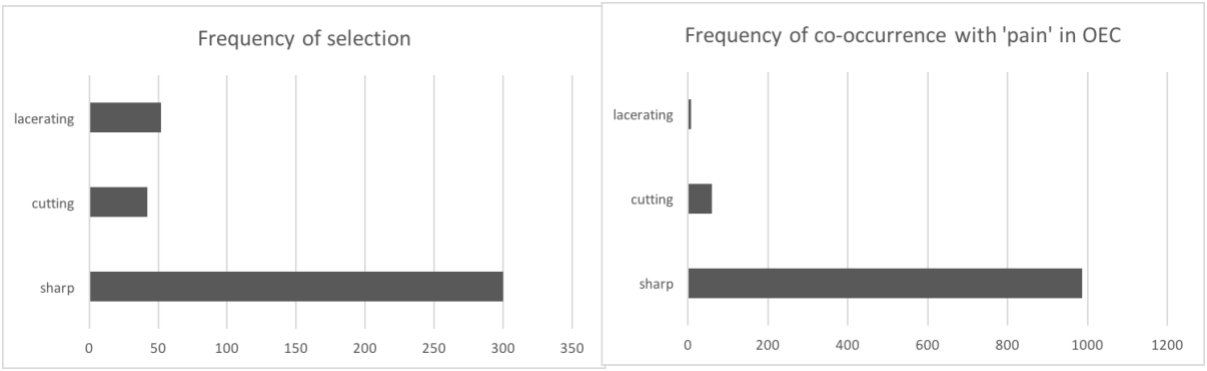


Figure 5 – Frequency of patient selections and frequency of co-occurrence with 'pain' in the OEC for Group 4 descriptors in the MPQ.



Group 6 and 7 exhibit the opposite pattern. As shown in Figures 7-8, both include a descriptor that collocates much *less* strongly with 'pain' than all the other descriptors in the group. Patients select that descriptor by far the least often. In group 6, this applies to 'tugging', which is at the bottom of the intensity scale; in group 7, it applies to 'scalding', which is in the top half of the scale.

Figure 6 – Frequency of patient selections and frequency of co-occurrence with 'pain' in the OEC for Group 6 descriptors in the MPQ.

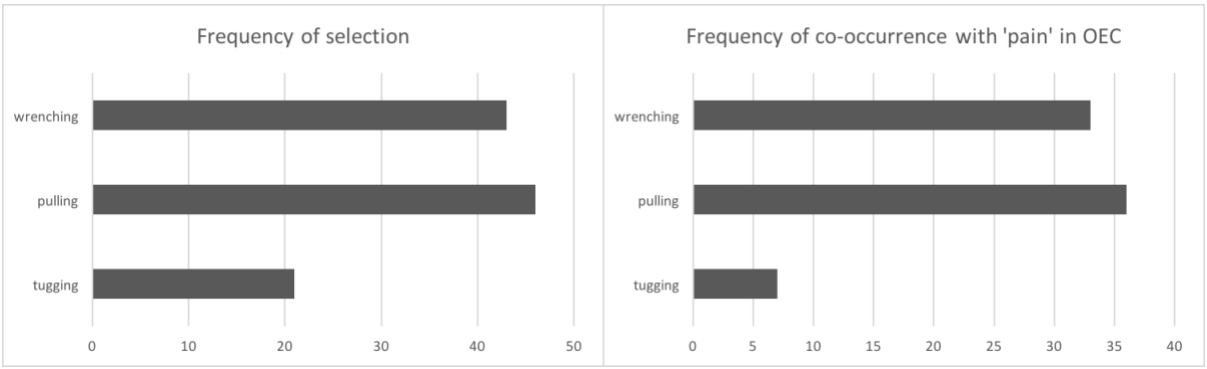
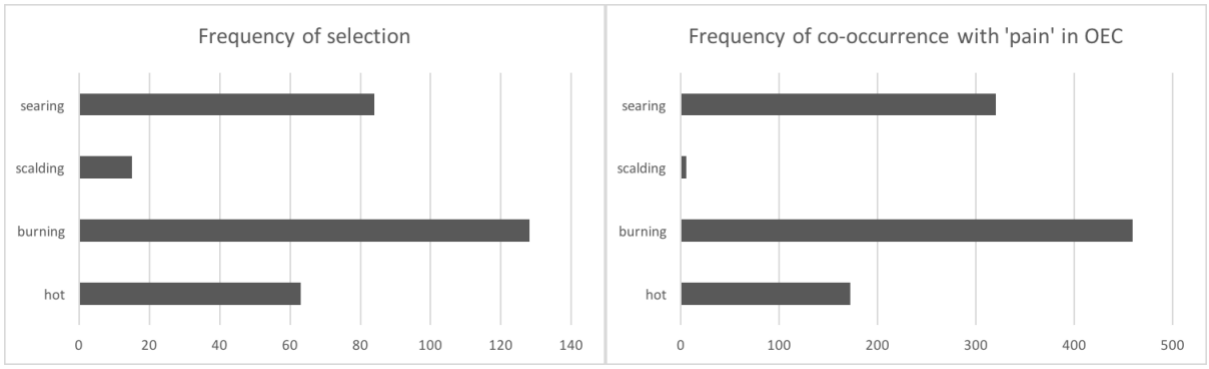
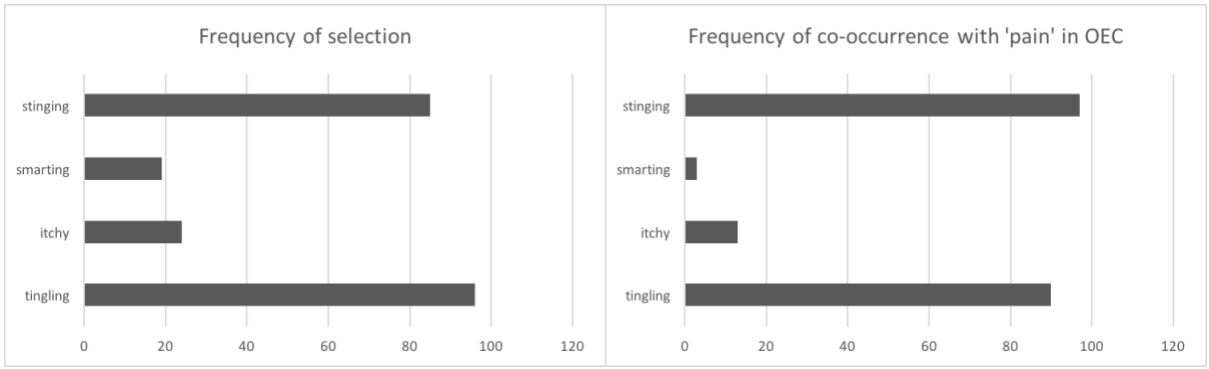


Figure 7 – Frequency of patient selections and frequency of co-occurrence with 'pain' in the OEC for Group 7 descriptors in the MPQ.



And finally, group 8 has two strong collocates and two weak collocates of ‘pain’ (‘tingling’ and ‘stinging’), as shown in Figure 8. Here too patients overwhelmingly select the two strong collocates, which fall at the opposite ends of the intensity scale: ‘tingling’ at the bottom and ‘stinging’ at the top.

Figure 8 – Frequency of patient selections and frequency of co-occurrence with ‘pain’ in the OEC for Group 8 descriptors in the MPQ.



In other words, for seven out of the 10 sensory groups in the MPQ, the choice of descriptor is explicable largely or entirely in terms of the strength of the collocational link from the word ‘pain’ to that descriptor. This is consistent with existing hypotheses and evidence concerning the psychological reality of collocational patterns, but undermines the validity of the intensity scales in these groups of the MPQ. If patient selections are driven by the priming effects resulting from different collocational strengths between the various descriptors and pain, they do not accurately reflect the severity of the person’s experience of their pain.

6. Implications for the use of the MPQ

The presence of a large number of linguistic descriptors is a strength of the MPQ, as they can open up opportunities for patients to reflect on and express the specific character of their pain, possibly for the first time. The affective groups, in particular, can encourage patients to discuss the emotional aspects of pain, which they may not have previously thought to be a relevant and appropriate topic in a medical consultation.

Nonetheless, as we have shown, the methods of corpus linguistics can provide an account of variation among the descriptors in the original MPQ that supports and further clarifies the concerns that have been raised in the clinical literature. Our analysis of the frequency and collocates of each descriptor in the OEC has identified the descriptors that are more likely to pose comprehension problems for patients due to lack of familiarity (i.e. those at the bottom of Table 1: 'tugging', 'quivering', 'tingling', 'pulsing', 'pinching', 'nauseating', 'gnawing', 'frightful', 'smarting', 'cramping', 'scalding', 'rasping', 'pricking', 'lacerating' and 'lancinating'), or cause confusion due to the prevalence of non-pain-related meanings (e.g. 'drawing' and 'boring').

In addition, we have shown the extent of variation in the strength of collocation between each descriptor and 'pain', and provided evidence of how this interferes with patients' selection in ways that undermine the ability of the MPQ to measure pain severity reliably, at least for some of the groups in the questionnaire.

This has implications, first of all, for the use of the original version of the MPQ, i.e. for the extent to which clinicians can rely on the scores that result from converting a patient's chosen descriptors into numbers. Patients' (lack of) familiarity with a particular descriptor, how tied a descriptor is to specific contexts of use, and the extent to which a descriptor is used in the context of pain can all influence these scores.

Second, our findings have implications for the process of selecting between the original MPQ and the two short-form versions. One of the innovations of the short-form MPQs is to measure intensity via numerical scales associated with each descriptor, rather than by attempting to list descriptors in order of intensity. While this approach makes the questionable assumption that all adjectives are gradable, it avoids the problem we have identified with the original version of the questionnaire.

Third, the collocation-related issue we have identified for the original English version of the MPQ is likely to apply to the many translations of the questionnaire into other languages, but the collocational patterns and contrasts within each group will be different for each language. As a result, any comparison of intensity scores across different language versions would be extremely problematic, or would minimally require an analysis like the one presented in this chapter for each translation.

7. Conclusions: Corpus linguistics and language-based diagnostic tools

Verbal descriptions are central to communication about symptoms such as pain in clinical settings. In this sense, the MPQ was rightly recognised as ground-breaking in going beyond numerical scales and including verbal descriptors that attempted to do justice to the rich and varied characteristics of patients' pain experiences. However, and inevitably at the time, descriptors were selected and grouped on the basis of a linguistically naïve view of contrasts between words, and without systematic evidence from actual language use.

According to Main (2016), this potential weakness applies more generally, however. In his 2016 systematic review of the literature on the MPQ, he comments that: 'the majority of existing pain quality measures have been developed primarily or solely based on expert opinion and often fail to include the descriptors patients report are the most important or common' (Main, 2016: 1389).

With the availability of corpus methods, there is neither need nor excuse to rely solely on expert opinion, however well founded, nor, for that matter, to ask patients

explicitly which descriptors they regard as most important or common. Rather, naturally occurring descriptions of pain (or of other symptoms or conditions) can be collected and studied systematically on a large scale. By relying on corpus linguistic evidence, future language-based diagnostic tools can avoid the pitfalls we have just described and include only lexical (and, ideally, also structural) choices that are likely to be appropriate for the largest possible number of patients.

Acknowledgement: The research reported in this chapter has been supported by the UK's Economic and Social Research Council (grant numbers: ES/K002155/1, ES/R008906/1).

References

Bennett, M. (2001), 'The LANSS Pain Scale: the Leeds assessment of neuropathic symptoms and signs', *Pain*, 92, 1-2: 147-57.

Brezina, V., & Gablasova, D. (2015), 'Is there a core general vocabulary? introducing the New General Service List', *Applied Linguistics*, 36(1), 1-22.

Brezina, V., McEnery, T., & Wattam, S. (2015), 'Collocations in context: a new perspective on collocation networks', *International Journal of Corpus Linguistics*, 20(2), 139-173.

Dworkin, R. H., Turk, D. C., Revicki, D. A. et al. (2009), 'Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2)', *Pain*, 144: 35-42.

Fernandez, E. and Boyle, G. J. (2002), 'Affective and Evaluative Descriptors of Pain in the McGill Pain Questionnaire: Reduction and Reorganization', *The Journal of Pain*, 3(1): 70-77.

Fernandez, E. and Towery, S. (1996), 'A parsimonious set of verbal descriptors of pain sensation derived from the McGill Pain Questionnaire', *Pain*, 66: 31-7.

Hoey, M. (2005), *Lexical priming: A new theory of words and language*, London: Routledge.

Hughes, J. & Hardie, A. (in press), 'Corpus Linguistics and Event-Related Potentials', in J. Egbert and P. Baker (eds.), *Triangulating Corpus Linguistics with other Linguistic Research Methods*, London: Routledge.

Kilgarriff, A., Baisa, V., Bušta, J. et al. (2014), 'The Sketch Engine: ten years on', *Lexicography*, 1: 7-36, 2014.

Lin, C-P., Kupper, A. E., Gammaitoni, A. R. et al. (2011), 'Frequency of chronic pain descriptors: Implications for assessment of pain quality', *European Journal of Pain*, 15: 628-33.

Main, C. J. (2016), 'Pain assessment in context: a state of the science review of the McGill pain questionnaire 40 years on', *Pain*, 157(7): 1387-99.

McEnery, T. and Hardie, A. (2012), *Corpus Linguistics: Theory, Methods and Practice*, Cambridge: Cambridge University Press.

Melzack R. (1975), 'The McGill Pain Questionnaire: Major properties and scoring methods', *Pain* 1:277-99.

Melzack, R. (1983), 'The McGill Pain Questionnaire', in R. Melzack (ed.). *Pain Measurement and Assessment*, New York: Raven Press, 41-7.

Melzack, R. (1987), 'The short-form McGill Pain Questionnaire', *Pain* 30: 191-7.

Melzack, R., Katz, J. and Jean, M. E. (1985), 'The role of compensation in chronic pain: analysis using a new method of scoring the McGill Pain Questionnaire', *Pain* 23(2): 101-12.

Ngamkham, S., Vincent, C., Finnegan, L., Holden, J. E., Wang, Z. J., Wilkie, D. J. (2012), 'The McGill Pain Questionnaire as a Multidimensional Measure in People with Cancer: An Integrative Review', *Pain Management Nursing*, 13(1): 27-51.

Rychlý, P. (2008), 'A Lexicographer-Friendly Association Score', *Proc. 2nd Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN*, 2: 6-9, 2008.

Semino, E. (2010), 'Descriptions of pain, metaphor and embodied simulation', *Metaphor and Symbol*, 25, 4, 205-26.

Wilkie, D. G., Huan, H-Y., Reilly, N. and Cain, K. C. (2001), 'Nociceptive and neuropathic pain in patients with lung cancer: A Comparison of Pain Quality Descriptors', *Journal of Pain and Symptom Management*, 22(5): 899-910.

APPENDIX

Table 3 – Collocates of MPQ descriptors in the OEC.

Group	Descriptor	Top 10 collocates in OEC	Occurrences of 'pain' lemma in +5/-5 window in OEC
4	Sharp	contrast, razor, knife, edges, decline, teeth, rise, wit, drop, edge	986
16	Intense	Scrutiny, pressure, heat, competition, lobbying, rivalry, speculation, pain, debate, fighting	809
9	Dull	dull, ache, boring, thud, lifeless, grey, senses, roar, grey, dreary	509
7	Burning	Fuel, fossil, fires, Man, coal, sensation, smell, candle, calories, flag	459
16	Unbearable	Lightness, stench, lightness, Kundera, pain, tension, Being, agony, heat, unbearable	458
19	Cold	War, weather, winter, hot, wet, water, cold, freezing, air, war	341
7	Searing	indictment, heat, pain, white-hot, riffs, temperatures, solos, honesty, guitars, exposé	320
18	Numb	Comfortably, anaesthetic, numb, fingers, senses, Floyd, toes, anesthetic, Pink, comfortably	295
2	Shooting	guard, drive-by, fatal, spree, rampage, shooting, gun, hunting, occurred, Columbine	251
1	Throbbing	gristle, headache, pulsating, cock, bass aching, pulsing, bassline, temples, ache	243
9	Sore	throat, throats, thumb, festering, cough, canker, muscles, loser, fever, swollen	210
3	Stabbing	fatal, shootings, fatally, Westward, beatings, knife, slashing, spate, knives, stabbing	200
20	Agonizing	slowness, reappraisal, wait, Mallman, 52.4, endured, agonizing, screams, indecision, dilemma	197
7	Hot	Topics, Copyright, copy, spots, water, cold, &, weather, hot, summer	172
9	Aching	joints, muscles, limbs, soothe, sore, throbbing, aching, legs, soreness, blisters	152
9	Heavy	rain, metal, rains, lifting, metals, traffic, burden, snow, rainfall, losses	142
17	Radiating	spokes, ridges, microtubules, Angelic, outward, acicular, outwards, concentric, T-Bone, warmth	135
14	Killing	wounding, civilians, injuring, innocent, accused, spree, convicted, unlawful, soldiers, Baghdad	133
5	Cramping	bloating, diarrhea, abdominal, Cramping, nausea,	118

		paresthesias, constipation, Abdominal, colicky, flatulence	
8	Stinging	rebuke, nettle, nettles, Nettle, tentacles, nematocysts, Urtica, dioica, jellyfish, rebukes	97
8	Tingling	numbness, sensation, sensations, spike, extremities, Numbness, spidey, itching, tastebuds, Spidey	90
18	Tight	Jeans, budgets, grip, controls, end, tighter, pants, budget, schedule, rain	90
20	Nagging	doubts, injuries, nagging, suspicion, groin, hamstring, feeling, injury, self-doubt, ankle	79
17	Spreading	virus, rumors, rumours, rapidly, wildfire, gospel, wings, worm, disease, flu	66
15	Blinding	effing, flash, randomisation, snowstorm, concealment, glare, randomization, Lights, flashes, sandstorm	65
5	Crushing	defeat, blow, defeats, dissent, disappointment, uprising, burden, inflicting, skulls, blows	63
4	Cutting	edge, taxes, grass, spending, cutting, pasting, costs, emissions, corner, Edge	60
9	Hurting	hurting, feelings, badly, sentiments, helping, knee, hurt, ankle, anybody, WHO	58
10	Tender	resignation, Sympathy, loving, mercies, meat, age, chicken, sympathy, juicy, simmer	57
18	Tearing	apart, shreds, ACL, ligament, ripping, ligaments, flesh, fabric, hair, cruciate	55
17	Piercing	tattoos, tattooing, tattoo, nipple, piercing, genital, armour, ear, gaze, navel	49
14	Cruel	inhuman, degrading, punishment, unusual, inhumane, Degrading, inhuman, Torture, torture	46
1	Pounding	drums, pavements, fists, surf, headaches, hooves, heart, pounding, fist, chest	43
19	Cool	pretty, cool, air, temperatures, breeze, stuff, weather, temperature, heat, warm	43
20	Dreadful	dreadful, penny, Penny, Dreadful, truly, mistake, appalling, tragedy, absolutely, ordeal	39
1	Beating	drums, heart, drum, brutal, torture, bush, Rodney, wings, severe, savage	37
6	Pulling	strings, plug, stops, trigger, hair, punches, out of, pushing, teeth, shirt	36
5	Gnawing	incisors, Glenfidich, fingernails, pang, gnawing, rodent, canker, unguardedly, emptiness, inside	35
6	Wrenching	gut, emotionally, dislocations, sobs, wrenching, Gut, heart, dextral, catharsis, adjustment	33
16	Miserable	miserable, wet, git, endure, wretched, bastard, lonely, failure, utterly, existence	32
16	Annoying	habit, annoying, incredibly, frustrating, pop-ups, ads, mildly, distracting, downright, fucking	31
18	Drawing	Paintings, painting, board, attention, room, prints, conclusions, photographs, pencil, ink	28

18	Squeezing	tightly, squeezing, pips, blades, margins, Squeezing, toothpaste, gently, sponge, lemon	28
14	Vicious	circle, cycle, Sid, spiral, unprovoked, assault, attack, Circle, virtuous, attacks	24
5	Pressing	button, buttons, ahead, issues, matters, flesh, lips, Closing, facing, keys	23
19	Freezing	Temperature, thawing, cold, sleet, rain, snow, freezing, temperature, below, assets	18
3	Boring	boring, dull, repetitive, bland, Boring, predictable, incredibly, tedious, mind-numbing, plain	15
10	Splitting	hairs, quadrupolar, headache, atom, quadrupole, hyperfine, deuteron, methyl, kHz, splitting	15
13	Terrifying	ordeal, prospect, conspiracies, subjected, terrifying, exhilarating, utterly, frankly, truly, relived	14
13	Fearful	fearful, anxious, Symmetries, Symmetry, reprisal, timid, backlash, lest, greedy, losing	14
16	Troublesome	groin, hamstring, knee, Reign, potentially, weeds, troublesome, Career, rid, ph	14
8	Itchy	Scratchy, rash, runny, sneezing, scaly, watery, scratchy, scalp, itchy, rashes	13
1	Pulsing	throbbing, veins, rhythm, throbs, pulsing, synths, synth, pulsar, techno, bassline	12
3	Lancinating	paroxysmal, burning, chronic, pain, or, is, as, of, the	11
14	Punishing	blocker, rewarding, schedule, perpetrators, punishing, deterring, offenders, criminals, sins, violators	11
15	Wretched	Fanon, momhood, wretched, Rabbit, Franz, miserable, villainy, teeming, dispossessed, hive	11
2	Flashing	lights, neon, sirens, siren, weep, weepers, strobe, Lights, amber, headlights	10
2	Jumping	bandwagon, bungee, conclusions, hoops, jacks, rope, queue, joy, jumping, fences	10
12	Sickening	thud, crunch, sickening, stench, lurch, disgusting, regularity, sameness, hypocrisy, Sickening	10
1	Quivering	jelly, lip, wrecks, wreck, jowls, blob, lips, goo, vibrato, nostrils	8
3	Pricking	pomposity, Whimper, Stockmarket, tingling, pinching, thorns, consciences, conscience, stinging, needle	7
4	Lacerating	razor-sharp, scorching, wit, rhythms, honesty, rage, portrait, critique, tongue, liver	7
6	Tugging	forelock, heartstrings, sleeves, forelocks, heart-strings, Tugging, trouser, leash, heartstring, tugging	7
17	Penetrating	sealers, gaze, sealer, armour, radar, stare, insight, rays, warhead, GPR	7
20	Nauseating	stench, pap, nauseating, sycophancy, hypocrisy, tosh, smugness, odors, repulsive, odour	7
7	Scalding	1-year-old, sulphurous, vents, hydrothermal, bath, gulped, hot, spew, tepid, burns	6

10	Taut	thriller, stressed, suspenseful, Bruno, rope, suspense, muscles, thrillers, abs, tummies	6
3	Drilling	offshore, ANWR, rig, arctic, rigs, wells, Wildlife, drilling, oil, Refuge	5
5	Pinching	pennies, penny, nipples, forefinger, nipple, pricking, sebum, chicks, multitouch, nock	5
13	Frightful	Hobgoblin, Fairytales, Fangoria, Obstacles, grimace, frightful, Ogilvy, howl, clutter, din	5
11	Exhausting	exhilarating, exhausting, emotionally, physically, stressful, mentally, avenues, draining, frustrating, remedies	4
20	Torturing	imprisoning, raping, murdering, detaining, prisoners, mutilating, maiming, executing, kidnapping, detainees	4
8	Smarting	meted, humiliation, defeats, thrashing, defeat, snub, debacle, humiliating, Hibs, acquittal	3
12	Suffocating	strangling, constrictions, oppressive, humidity, claustrophobic, possessiveness, choking, Sheberghan, stench	3
14	Gruelling	trek, chemotherapy, schedule, marathon, mile, slog, triathlon, Peaks, marathons, endurance	3
1	Flickering	candles, candlelight, candle, flame, firelight, flames, torches, lights, shadows, glow	2
11	Tiring	visibly, stressful, tiring, time-consuming, draining, exhausting, tedious, dreadful, physically, tanka	2
10	Rasping	forehand, cross-cut, left-foot, radula, 25-yard, cross-shot, half-volley, cannoned, breaths, guttural	0

[i] Group 4 is also one of several groups in the MPQ where a particular quality of pain sensation is captured via metaphorical references to causes of physical damage to the body (see Semino, 2010).

[ii] The computing of collocates via Sketch Engine was case-sensitive, which means that, in some cases, the same collocate appears twice in Table 3 for the same descriptor, with and without an initial capital.

[iii] In a few cases, some of the top 10 collocates relate to a widely quoted title or line (as with Milan Kundera's *The Unbearable Lightness of Being* for *unbearable* and Pink Floyd's 'Comfortably numb' for *numb*) or widely reported events (as with the Columbine school massacre for *shooting*).