

**Subtext as humour:  
Grammatical 'hidden meaning' as part of pun and metaphor in English**

Marija Milojkovic

Department of English, Faculty of Philology, University of Belgrade

marija.okon@gmail.com

*Павлу Михайловичу Кожину,  
который, к моему изумлению,  
все понял<sup>1</sup>*

**Abstract**

Corpus-derived subtext is a new development in Contextual Prosodic Theory (Louw 2010a), which this paper employs, among other tools provided by CPT, in the attempt to give corpus-attested definitions of pun and metaphor. Subtext is defined as the most frequent lexical variables contained in a given grammatical string. Corpus stylistic research (Louw and Milojkovic, 2014) has shown that subtext participates in conveying meaning, albeit at a subconscious level, because lexical variables within grammar strings are opaque to intuition. In the process of providing the definitions of these two vastly different literary devices it becomes apparent that, although there is no grammar without its most frequent lexical variables, these do not participate in meaning creation when it comes to pun if the purpose of this device is solely to amuse, as in comic one-liners. When pun is part of a grander design, however, within a broader context of situation, and where both its meanings create their own states of affairs participating in the transition in a text, corpus-derived subtext does play a meaningful role. Metaphor, unlike simple wordplay, always involves subtext, which can be objectively revealed by studying its interaction with the text's context clues.

**Key words**

corpus stylistics, Louw, collocation, semantic prosody, corpus-derived subtext, pun, metaphor.

---

<sup>1</sup> This paper is dedicated to Pavel M. Kozhin (1934-2016), Institute of Far Eastern Studies, Russian Academy of Sciences, in acknowledgement of his academic support, which began at the time of my writing this paper.

## 1. Preliminaries

### 1.1 Collocation and states of affairs

What does humour consist of? Ostensibly, it is a message in context that causes amusement. Hardly anyone would doubt the importance of these three preliminaries: a meaning of a message is always constructed and re-constructed in context, and, to be humorous, it must cause laughter, even if cynical, bitter or vacuous. It goes without saying that a message need not be verbally expressed to have meaning. If, however, it uses language as its medium, the basic premise is that we are dealing with a stretch of text that someone (not necessarily both its author and its recipient) will think funny.

Several of these premises need further defending and, possibly, further expanding. On the one hand, even a meaningless message may be considered funny precisely of its lack of meaning. Then, a message containing meaning may be thought funny by its creator but not by its addressee (there have been instances of public shaming on Twitter for messages which their senders considered jocular and inoffensive). Finally, a message may amuse without being intended to. This last quality is often explored in novels, when a literary character is being funny in the eyes of the reader, but not in his own or those of his audience.

All these cases merely confirm the importance of context. A meaningless message is amusing in a context that requires it to be meaningful. A mishap on Twitter means that the public views the message in a different context from its author. An amusing statement coming from a speaker who remains in the dark as to the effect of her words again suggests that the hearer views it in a different light.

After all, there are theories emphasising that language as such has no meaning unless there is a mind understanding it in context. They go so far as to say that language is a meaningless stretch of signs before a mind finds a context in which this stretch of signs starts to make sense ('The meaning is not in the words; it is created in the mind of the addressee' (LaPolla, 2006). This is both true and not true. From the point of view of the conscious individual receiving the message, it is. But language, even if viewed as mere signs as yet uncomprehended by a mind, must have been produced by another mind and a language carrier. Jotted in pencil or engraved in stone, it is an instance of language, created in context with a message in mind. No matter whether the message is understood as it was meant to be by its author, or whether it is

received in a context that imposes additional layers of meaning, if one person is amused by it, it is enough for it to be considered funny.

The complexities involved in defining, describing and explaining humour have been researched in a body of works (see Chiaro, 1992, Simpson, 2003, Goatley, 2013). Still, there seems to be a relative paucity of corpus-based approaches that would explain and define the creation of humour purely in corpus terms. A successful attempt is made by Louw (2014), in which humour is viewed through states of affairs, created in the reference corpus by co-selection. The idea of the paper (also a keynote at the 2009 Lucca Conference on Humour Studies) is that if two words used in relative proximity (within John Sinclair's nine-word window) are found to create a certain state of affairs in the reference corpus, there is a 'joke waiting to happen'. For example, 'back' co-selected with 'stool' will often be shown to participate in contexts where a person falls off a stool on their back, or because a stool does not have one. The potential of language to create humour does not reside only in co-selection, which acts as a situational prompt creating states of affairs in the reference corpus that can be manifestly similar situationally. 'Irony in the text or insincerity in the writer' (Louw, 1993) is focused on a more verbal and less situational means to conjure up humorous associations. According to Louw, a writer/speaker using an unexpected collocate may thereby awaken different associations in the reader/hearer. They may not be aware of the particular mechanism of how a humorous impression is formed (the semantic auras of two collocates 'clashing' in a context which ostensibly endorses one implication and covertly promotes another), but their accumulated language experience steers them towards the humorous effect nonetheless. The reference corpus is the analyst's tool, helping scientifically to show what particular associations come clashing together because of the writers'/speaker's choice, and precisely how the context of situation ensures that the reaction is laughter rather than indignation or bafflement.

Co-selection is the connection between purely situational humour (which may not involve language) and humour as verbal art. If we place Mr Bean in a room with a stool, it is unlikely that he will simply sit on it (of course, the comic potential of this arrangement may defy the audience's expectations). To prove the existence of this comic potential, we have a whole corpus of Mr Bean episodes and sketches to refer to. In a similar way, if we find two words in reasonable proximity to one another in a stretch of text, the reference corpus may show that this proximity may be humorous. We may also find that together they create a particular state of

affairs in the corpus, which may be infringed by introducing collocates that are incongruent in a funny way (the incongruence will also be shown in the corpus). The reference corpus acts as a sample of the world, and word frequencies in the corpus are not more important than frequencies of situations in which they participate. A word on its own carries no meaning unless it relexicalises (i.e. attains its full lexical meaning) in the mind of our language carrier. For example, what does 'decompose' mean on its own? Only in this is LaPolla (2006) correct: a word on its own, rather than a stretch of text, carries no meaning unless the mind encountering it creates a context that relexicalises it. Of course, if the context is already there to help, it is part of the message (e.g. a pair of shoes in a shop window accompanied by a price tag). But as for frequency lists, they should not be taken at face value unless they are accompanied by contexts.

So far, we have established the analyst's tools that, according to Louw, scientifically, i.e. with the help of a reference corpus, explain the effect of verbal humour. One is co-selection: two or several words get co-selected in a text, and in the reference corpus they chunk a certain state of affairs. Either the set-up is humorous to start with, or it will become so by introducing a third collocate. Imagine a room with a judge and a jury and then introduce into it an incongruent element, like a carrot, or Mr Bean. Our experience of life (for which a reference corpus is a verbal substitute) does not allow carrots in courtrooms. The analogy is not entirely misplaced simply because humour is often purposefully created. The other satirist's/analyst's tool is exploiting/exploring semantic auras. Louw (1993) establishes them scientifically as semantic prosodies, discovering that many words, maybe the majority, are used in the language almost exclusively in a certain way. For instance, if a word is used mainly negatively, and the writer uses it in a seemingly positive context, the effect is ironic. This is the case with the conference goers who are 'bent on self-improvement' (Lodge, 1984). It turns out that 'bent on' is often followed by negative collocates, like 'destruction'. Semantic prosodies are different from connotations and associations because they are sometimes impossible to establish unless through a reference corpus. Many were surprised to find that 'happen' has a negative semantic prosody (Sinclair, 2003: 117).

The reference corpus, then, is viewed by Louw as a tool that acts both as a sample of the world and as a proxy for the accumulated language experience of a language user. Of course, language experience will differ from user to user, hence the existing theory of priming (Hoey, 2005). Going back to LaPolla (2006), a mind encountering a stretch of text and finding a context

in which it makes sense will apply to it only the experience, linguistic and otherwise, that the mind possesses. Much will depend on the person's education, background, psychological type, present mood and even political beliefs (if you google Oksana Boiko, RT's journalist, you will discover that she is either rude and aggressive, or honest and professional, never mind that these are reactions to the same interview!). If the individual mind brings so much baggage to the stretch of text they are about to process, composed by another individual with perhaps similar language experience but a necessarily different outlook, it is a wonder that the message is not doomed. At the other end of the spectrum there are spouses who understand one another using very few words or no words at all. It is clear that contextualising is the key. However, the reference corpus shows enough of consistent usages to confirm our everyday impression that, despite the hazards of individual contextualisation, the majority of us manage to get along reasonably well.

What exactly are these consistent usages that the corpus shows to exist? Its capacity is far greater than pointing to morpho-syntactic laws and pragmatic conventions. As shown by the existence of semantic prosody (SP), words, expressions and whole chunks of text may possess semantic auras that we cannot intuitively fathom (as shown by the SP of 'happen'). The point is not that some SPs are not wholly opaque, but that we can use a reference corpus on each occasion to establish exactly how an expression is used. Nor is this phenomenon restricted to lexis. When Louw (2014 [2009]) mentions 'get\* \* hands on' referring mostly to another person's money or body or both, he comes close to his own discovery (see Louw, 2010a and onwards): not only lexis, and not only lexico-grammatical combinations can have semantic auras. Pure grammar has auras also. Louw discovered that grammatical strings have frequent lexical collocates that determine these strings' semantic auras in cases where the lexis of the author's choice is less than frequent. Louw called this type of semantic prosody 'logical semantic prosody – subtext'. The mechanism of corpus-derived subtext will be explained in the next section.

## 1.2 Semantic prosody and subtext

Semantic prosody, then, is distinguishable from connotation on the grounds of its intuitive opacity, on the one hand, and scientific precision, on the other. It does not yield itself to intuitive analysis, unlike connotation, and is precisely established through its contexts in a

reference corpus (we do not speak of precision when discussing connotation). Although we as language users are not aware of SP, we still generally tend to produce language that confirms its existence (for example, 'give rise to' tends to be negative when it is not neutral (Xiao & McEnery, 2006)). Foreigners, or children, whose accumulated experience of a language is still insufficient to take account of SP, produce text that breaches it. For example, a foreign student may thank their supervisor for their 'persistent' help and advice (Hunston, 2007: 259). SP is therefore more subtle than connotation, and has more to do with currently expressed attitudes than with general opinions. For example, 'a mercenary' carries negative connotation, as opposed to the quality of being 'persistent', whose SP (or the fact of breaching it) is determined by its collocate. This is what Louw means when stating that SP is 'more strictly attitudinal' than connotation. Perhaps a better phrasing would be 'more finely attitudinal'. A good example of connotation is the difference between 'terrorists' and 'rebels' in the world of newscasting, which has more to do with the affiliation of the broadcasting house and less with the actual misdeeds of what may turn out to be one and the same organisation. A good example of SP is the already mentioned 'give rise to', the only example of which in the Cambridge Dictionaries Online (accessed on 29 November 2015) contains only positive collocates (<http://dictionary.cambridge.org/dictionary/english/give-rise-to>).

According to Louw (1993), in cases where SP is breached by native speakers, two out of three such cases will be instances of irony (such as 'bent on self-improvement) and one of insincerity. Insincerity here is not equated with an intention to deceive, but rather to conceal one's true attitude – for example, not to hurt another person's feelings. We are all familiar with the sort of situation when we feel ill at ease and the phrasings we come out with become slightly unnatural. Insincerity at this juncture covered the whole scope of expressions that contains treacherous SPs betraying a lack of frankness, from superficial everyday politeness to downright deliberate lying. It was because Louw's theory at that stage was at the level of the lexical, and we can always somehow account for the lexical, using hindsight. When it comes to grammar, which is completely opaque to intuition, the issue surpasses that of a definable message that somehow got mutilated in conveying, and instead we look at an additional layer of meaning. The reason for this is complete opacity: any phrasing must convey a grammar string, and any string will carry subtext. If we only used grammar strings with their most frequent lexis, and thus avoided deviation, language use would turn into a string of clichés. On the other hand, a lexical collocate

so uncharacteristic as to be intuitively spotted as such will most probably also be ungrammatical, as grammar reduces the scope of collocating lexis.

Here we come to the main difference between SP and corpus-derived subtext: the propensity for deviation. For example, we know that ‘be cruel to be kind’ is a paradox, because the meaning of one excludes that of the other. This is a conflict of connotations. ‘Bent on self-improvement’ is a similar semantic relationship, but at the level of SP. We feel it too, only more subtly. It does not stand alone, but is embedded in a stretch of text containing other context clues from which we infer the text’s meaning. But the semantic aura of a string consisting exclusively of grammar words cannot have a meaning to us. Therefore, lexis either fits into it or it breaches a morpho-syntactic rule. Grammatical auras will always betray additional meanings, unless the string only takes one lexical collocate. For example, in the Google Books corpora (Davies, 2011), the grammar string ‘is more \* after than’ yields ‘sought’ as the only lexical variable (Milojkovic & Louw, 2017).

Inscrutable as it is, can grammar express irony? In the absence of other helpful context clues, which must be lexical, I at present believe it cannot. But it must add a layer of meaning that we feel subliminally, and which must explain the cases when we feel that a certain writer’s style is ‘better’ than another’s for reasons that escape us.

Several examples of humorous subtext in Lodge have already been found (Milojkovic, 2013; Louw & Milojkovic, 2016). They show that insincerity comes in many guises, including wishful thinking and a desire to conceal a love affair. The inscrutability of grammar ensures that where we have successfully fooled ourselves or others lexically, we will not succeed in doing so lexico-grammatically. Grammar seems always to speak the truth. The source of humour will then lie in our pathetic attempt, rather than in incongruity discussed in Section 1.1. Grammar strings do not accept incongruous collocates, even those that reveal themselves as such with hindsight. Humour, on the other hand, by definition requires incongruity that can be registered without recourse to reference corpora. Simply put, humour needs to be understood. Still, the subliminal message of grammar strings is not to be ignored, now that we have the means as well as the end: to access the full meaning of any message. So what if the only person laughing after accessing corpus-derived subtext will be the corpus analyst? He who laughs last, laughs longest... and laughs him/herself into relexicalisation.

### 1.3 Relexicalisation and literary devices

Relexicalisation was mentioned in Section 1.1 in the discussion on how a word without any context at all attains its full lexical meaning in the mind of the language carrier if they create a context around it in the mind. For example, seeing 'take' may bring to mind the act of taking an object with one's hand, and not expressions like 'take a seat' or 'take a look'. The reference corpus, on the other hand, informs us that 'take' as part of the latter expressions is more frequent in the language than in its full lexical meaning, 'take an object using one's hand'. Sinclair's term for the meaning of 'take' in 'take a look' was *delexical* (Sinclair, 1987), as opposed to being fully lexical (literal).

However, in the last sentence of the previous section, the idiom 'he who laughs last, laughs best' takes on two meanings simultaneously: the delexical and the literal. It so happens that the corpus analyst 'laughs last', because he or she takes longest to perform the analysis. But their appreciation of the humorous potential of the given turn of phrase will be the most complete, and therefore they will 'laugh longest'. This situation-dependent meaning had to do with the timing and the quality of laughter and is therefore literal. But the delexical meaning of the idiom as it is used in English also applies: it is the final result that counts most. This effect of double meaning is achieved because, in the context of humour and subtext, the verb 'laugh' is understood literally. It regains its literal meaning as part of an expression that is otherwise used delexically. In fact, this double meaning results in wordplay, or pun.

Louw (2008) claims that relexicalisation is the basis of all literary devices (of which wordplay is, arguably, one of the lowly members). Moreover, he calls for creating a glossary of corpus-attested literary terms. The main idea is that our intuitive judgement of whether a turn of phrase is wordplay, metaphor or irony should become scientific. As examples, Louw lists his own proposed definitions of semantic prosody, irony, insincerity, and metaphor.

Louw's claim that relexicalisation has a scientific status rests on his more general claim that collocation has a scientific status. The example with the idiom 'he who laughs last, laughs best' relexicalising in the context of humour is mainly based on situational prompts and the fact that the verb 'laugh' is used in its literal meaning in the immediate proximity of the idiom. In fact, relexicalisation is thought by Louw to work mainly colocationally rather than situationally. Let us take Yeats's 'The Circus Animals' Desertion':



Summer and winter, till old age began,  
My circus animals were all on show...

The meaning of 'on show' is undoubtedly delexical. 'Circus' and 'show' are separated by four words, which fits into Sinclair's nine-word window of collocative power (four words to the left of the node and four words to the right). This means that 'circus' and 'show' exercise collocative power over one another. If we co-select these two words in the British National Corpus (BNC), under the same conditions (4 words at most between them), we will find that they all chunk the state of affairs of a circus show (the reader is referred to Davies (2008-)).

As previously stated, Contextual Prosodic Theory treats the reference corpus as a sample of the world. Simultaneously, and closer to home, the reference corpus is taken by it to represent the language norm. Since the reference corpus is balanced and representative, and thus reflects the language it is a sample of, we may assume that it more or less represents the accumulated language experience of its speakers. This is confirmed by tendencies in language use that are consistent but cannot be intuitively accessed, such as SP and subtext. Therefore, if 'show' relexicalises in the presence of 'circus' in the reference corpus, it must relexicalise for the native speaker. As it was shown in the paragraphs above, it can be a basis of humour in the cases of wordplay. Since Yeats's 'my circus animals were all on show' does contain wordplay, it may not be amiss to conclude that this line also contains an element of humour, despite the melancholy tone. If that is so, then the humour must be directed at the 'animals', or literary tricks and devices, and it could be surreptitiously subjecting them to ridicule: not only have they abandoned the elderly poet, but perhaps they were less than adequate in the first place? Hence their being equated with what ostensibly belongs to the circus? Before drawing the final conclusion about this line in particular, let us dwell on the mechanism of wordplay, or pun.

## **2. Defining wordplay**

These two definitions of pun can be found in traditional sources: 'a figure of speech which involves a play upon words' (Cudden, 1999), and 'a play on words that are either identical in sound (homonyms) or very similar in sound, but are sharply diverse in significance; an example is the last word in the title of Oscar Wilde's comedy, *The Importance of Being*

Earnest' (Abrams and Harpham, 2009). Despite the difference between these definitions, both sources give the following two examples, among others: John Donne's 'Hymn to God the Father' ('And having done that, Thou hast done'), and William Shakespeare's *Romeo and Juliet* (Mercutio: 'Ask for me tomorrow, and you shall find me a grave man'). In the first case, the pun is sustained through the homophony of 'done' in the meaning of 'finished' and 'Donne' as in John Donne: God will accept Donne's soul after he has forgiven all his sins. In the second example, the dying Mercutio plays with the meaning of 'grave' as a noun and 'grave' as an adjective, while the immediate co-text of the pun contains discussion of his mortal wound.

The following example of wordplay is from a novel by Charles Dickens (1993: 368):

'I say, old boy, **where do you hang out?**'

Mr Pickwick replied that he was at present **suspended** at the George and Vulture.

In this context, 'where do you hang out' is a fixed expression, meaning 'Where are you to be found?' The word 'suspended' is never used in this context, but it does reactivate the literal meaning of 'hang out', as in 'to hang out the washing' (Cobuild 1998).

An example of pun that in at the same time a paradox can be found in Philip Larkin's 'Winter Nocturne'. The poet describes the advent of night as unrest in nature. The poem finishes with

The rain falls still: bowing, the woods bemoan;

Dark night creeps in, and **leaves the world alone.**

The whole poem is about the period of transition between day and night, culminating in its arrival. How then can night at the very moment of arrival leave the world alone? Apart from the meaning of 'leave something in a particular state' (which implies that the world is alone now night has entered it), Larkin must also have had in mind the meaning 'not bother or interfere

with, leave be'. The word 'leave' co-selected with 'alone' was searched in the 1995 Times newspaper corpus. A hundred and fifty concordance lines were found. In fifteen of them (10%) the context was literal (e.g. leave a child alone, i.e. without adult supervision). Out of the remaining 135 delexical usages, 99 (73%) pointed to the conclusion that leaving something or someone alone in the sense of not interfering with it is a good idea, because of the unwelcome consequences of not leaving alone. Therefore the poem about the advent of night describing unrest felt in nature prior to its arrival states clearly that night, by leaving the world in the state of loneliness, does a good thing (Milojkovic, 2011).

Puns are often the basis of comic one-liners, e.g. 'Santa's helpers are called **subordinate clauses**'. Here 'subordinate clauses', a frequent collocation in the reference corpora, becomes viewed as a combination of 'subordinate' in the sense of 'less important' (Cobuild 1998) and 'clauses' in the sense of 'little santas'. Puns also often feature in advertisements, or literary titles – for example, *Brothers in Law* is a comic BBC serial about an idealistic young lawyer (<http://www.bbc.co.uk/programmes/b050gvnz>, accessed on 29 November 2015).

If we look at these examples from different genres, it will be clear that the majority contains a frequent collocation, one of whose constituents undergoes relexicalisation. Therefore, the following corpus-attested definition emerges:

Pun is characterised by a double meaning of a lexical collocates which is part of a frequent and often delexical collocation, whether lexical ('grave man') or lexico-grammatical ('thou hast done'). The double meaning is composed of the meaning of the lexical item as part of the frequent and delexical collocation, as found in the reference corpora, and a provisional meaning, often literal, imposed on it by context clues, whether immediate ('Santa's helpers are known as subordinate clauses'), distant (the discussion of Mercutio's wound) or extra-textual (the surname Donne is not in the poem where the pun appears). Relexicalisation (reducing a more frequent and delexical meaning to a less frequent and literal one) may be absent in cases of homophony (done/Donne), or homonymy (the two meanings of 'grave' are not etymologically related).

This definition should be checked, and probably refined, against other existing examples of pun. At this juncture, however, it has already become obvious that collocation, as understood by Sinclair (as co-occurrence within a certain distance in the text, the distance being measured by words) does not wholly determine this literary device, at least not on all occasions. 'Santa's helpers are called subordinate clauses' is a perfect example of Sinclair's nine-word window, with exactly 4 words intervening between 'Santa's' and 'clauses'. But that does not have to be the case: in some cases it is necessary to have the knowledge of the macro-context, and even external factors (the author's name, for example) in order to appreciate to the full the effect of wordplay.

This definition has been arrived at in the top-down way, and seems to work in cases of what I *know* to be wordplay. It is to be hoped that it also works bottom up, and an independent entity not possessing intuition, such as the computer, could highlight the cases of wordplay according to the definition's instructions. But what is the actual source of humour? When is the computer to laugh (let alone smile)? It appears that, since the provisional, literal meaning is the new element in the arrangement, whereas the delexical meaning is the one normally used in the language and found in the corpus, it is the provisional meaning that provokes laughter.

On the basis of these examples, the relationship between wordplay and humour is not simple. Leaving aside the assumption that a sense of humour varies from person to person, we may assume that one-liners found on comic websites are relied upon to be found funny by the majority of native speakers of that language. Their topics are universal, because they must be if they aim at the average reader. Some jokes, however, are only funny if you are familiar with their social or political context, or if you hold certain political beliefs, and this is when Malinowskian context of culture comes to mind. But solely on the basis of these examples, we can distinguish between serious ones (Donne, Larkin) and humorous one (Dickens, the BBC series title). It appears that the impression of incongruity prompts laughter, while if the collocation simply fits the context in two different ways, the reader is appreciative rather than amused. At the same time, the point of a one-liner does not go beyond the laughter-inducing quip. 'Subordinate clauses' are their own point, there is no broader context where they play a part if literally taken. Judging by Donne and Larkin, double meaning in which both senses contribute to the message seems to emerge precisely in cases where laughter is not the immediate

goal, if not expressly out of place. But even in the latter cases, one feels that if the author has risen to wordplay, the tragedy is not absolute.

The computer, then, may be taught not to laugh if it registers certain context clues that preclude laughter. The point of this lengthy discussion, some of whose premises have already been stated elsewhere, is that figures of speech may far overstep the limits of their immediate contexts.

I have already expressed the opinion that subtext cannot have humorous implications, unless very subliminally. In fact, it may have a completely different semantic aura and still not interfere with the jocular implication. There is no mention of subtext in the offered definition of wordplay. There is a reason for this: although examples of wordplay, even if one-liners, must contain lexico-grammatical collocatons and therefore subtext, the subtext of the grammar strings in question can never be a rightful participant in creating puns provided they are their own goal. An example of subtext in puns is discussed in Louw & Milojkovic (2016). The joke in question is Hot Tomato:

Q: Why did the Tomato blush?

A: Because he saw the salad dressing.

As it turns out, the grammar string 'he \* the \* \*ing' contains a frequent lexical item, or quasi-propositional variable, 'saw'. However, the semantic aura of the string clashes with the humour genre, because it appears either in religious contexts, or in those describing a (dangerous) turning point (Louw & Milojkovic, 2016). It may be that such contexts of situation always inspire more consideration than, on the face of it, is afforded them by their observers or even participants (ibid.). Our concern at present is to state that, although subtext is always present in a lexico-grammatical collocation, it cannot have a point in 'word-play', as the very term suggests. Puns serve a different purpose altogether, their conspicuousness, purely lexical, being both their means and their end. The purpose of pun is not to convey meaning, it is to convey double meaning (as in the quote by John Donne, where both meanings intertwine to stunning effect). I speculate that in the cases where the intention is to amuse, rather than to convey meaning, the attention is so much at the level of the lexical that the subtext of the emerging grammar string,

while it is being ‘fit’ into the lexical arrangement, becomes slightly artificial. That is why subtext is not mentioned in the definition of wordplay, offered above.

The case of John Donne supports this assumption, in that his wordplay, perhaps not meant to amuse, but genuinely to arrive at two meanings that both fit the context, does contain subtextual meaning. ‘Thou hast done’ may be viewed as a grammatical string. It is contained in four contexts in the BNC (one of them negative and none religious), and in six contexts in COCA (half of them positive and could be described as religious, the others not religious and negative). If we go to the Google Books – UK corpus, however, we first encounter biblical contexts, which is appreciated because of the similarity of the context of situation. It turns out that ‘thou hast done’ can have a positive or negative meaning, depending on who is uttering the words. In the first twenty contexts, if God addresses man (or woman), they are (in these contexts) all negative, blaming mankind for various sins (unlike the context in COCA where God speaks to Abraham: ‘...because thou hast done this thing, and hast not withheld thy son, thy only son, that in blessing I will bless thee...’) If it is man who speaks to God, then, naturally, or at least in the contexts studied, we witness only praise and trust in God’s judgement. Thus, subtext in John Donne is in accordance with its contexts of situation, and, even if the roles of the speaker and hearer are reversed, it is human sins that come into focus, just like in Donne.

Larkin’s ‘and leaves the world alone’ may serve to verify this assumption. It is perhaps less amenable to analysis, but at least the semantic aura is confirmed by the BNC. The line ‘and \*s the \*’ yielded the following concordance (‘stray’ lines, containing grammar words, as well as those not containing verbs and nouns where they are found in Larkin’s line, and variants inappropriate for other reasons were excluded):

- 1 and has the power 12
- 2 and has the advantage 10
- 3 and takes the form 9
- 4 and opens the door 7
- 5 and forms the basis 6
- 6 and gives the impression 6
- 7 and has the potential 6
- 8 and reserves the right 5

9 and shuts the door 5

10 and concerns the land 4

11 and includes the PAT 4

12 and reduces the risk 4

13 and tells the story 4

The most frequent QPVs, ‘power’ and ‘advantage’ do correspond to Larkin’s context, as the night is viewed both as powerful and an advantage from the point of view of the persona. Also, lines such as ‘has the potential’ and ‘reduces the risk’ in the concordance point to a positive semantic aura, although, of course, other contexts can be negative as well (for example, ‘opens the door’ is positive only in three contexts out of seven, and serves as a transition to a dangerous situation). Still, Larkin, like Donne, uses subtext that fits the context of situation, despite the existence of the pun in the final couplet. The difference in the behaviour of subtext in these two poetic contexts and in the two-liner ‘Hot Potato’ is enormous: it deepens the meaning in the former two cases and counteracts it in the latter case.

### **3. Defining metaphor**

In this article, the mechanism of relexicalisation was explained, using the example of Yeats’s line from the poem ‘My Circus Animals’ Desertion’. In order to proceed and explain the metaphor inherent in that particular line, it will be useful to look at the whole first stanza:

#### I

I sought a theme and sought for it in vain,  
I sought it daily for six weeks or so.  
Maybe at last, being but a broken man,  
I must be satisfied with my heart, although  
Winter and summer till old age began

**My circus animals were all on show,**

Those stilted boys, that burnished chariot,

Lion and woman and the Lord knows what.

Relexicalisation in the highlighted line is evident, and has been discussed. Also, intuitively we conclude that ‘circus animals’ must mean ‘literary devices’, or something to that effect: they are the ‘tenor’, or ‘target’, and ‘circus animals’ are the ‘vehicle’, or ‘source’. If we look for a corpus-attested confirmation of this hunch, we can co-select ‘device\*’ with ‘theme’ (‘theme’ being the issue under discussion in the stanza) and find the following single context in the BNC: ‘Eikhenbom’s essay on Gogol’s Overcoat is a good example of the latter. He sets out to show that in Gogol’s tale the centre of gravity is transferred from the **theme**... to the **devices** (1963: 377)’. The COCA (Davies 2008-) yielded five contexts, in which those of artistic creation predominated:

1 what I call the ‘shaggy-dog story’ **device**. The opening **theme** of the Eroica be  
2 ariation on the solar **theme** is the **device** developed in Florida Solar Energy c  
3 family (of TV and movie fame) as a **theme**. The complex **device** ended when Thing  
4 Wood’s use of the seasonal **theme** as a **device** for expressing the doctrines of  
5 Besides these overt lexical **devices**, the **theme** on which the plot hinges, enco

Co-selection chunks the states of affairs of composing music (line 1), technical inventions (lines 2 and 3), painting (line 4) and lexical devices in American SF (line 5). Thus, it is possible to establish the implied subject of the metaphor through co-selection and the states of affairs in the corpus.

The contribution of this article to (corpus) stylistics is the discovery that corpus-derived subtext generally seems to play a major role in Yeats’s metaphors, albeit invisibly. The descending frequency list of the items in the lexical slot of the line ‘my \* were all’ in the Google Books-UK corpus is found below:



my **friends** were all (156)  
my **thoughts** were all (136)  
my **children** were all (64)  
my **companions** were all (60)  
my **eyes** were all (59)  
my **nerves** were all (56)  
my **men** were all (53)  
my **clothes** were all (49)  
my **sympathies** were all (46)  
my **family** were all (45)  
my **hands** were all (41)  
my **dreams** were all (40)

The QPVs of 'my \* were all' can be divided into four categories:

<b>I</b>	<b>II</b>	<b>III</b>	<b>IV</b>
<b>friends</b> (156)			
	<b>thoughts</b> (136)		
<b>children</b> (64)			
<b>companions</b> (60)			
		<b>eyes</b> (59)	
	<b>nerves</b> (56)		
<b>men</b> (53)			
			<b>clothes</b> (49)
	<b>sympathies</b> (46)		
<b>family</b> (45)			
		<b>hands</b> (41)	
	<b>dreams</b> (40)		

Judging by the wider contexts from the reference corpus, the first category can be described as 'significant people in one's life, such as family and friends'. The second is 'thoughts and emotions' (with one exception as far as I could see, that is, the meaning of 'nerves' in the reference corpus). The third is significant because these organs we need to see and do things. The fourth, 'clothes', is in the reference corpus often in bad condition, with the implication that it may endanger the health of the narrator. While all the four categories are significant, the first two have a direct bearing on the poem.

Let us start with the second category. In order to illustrate the presence in the poem of this subtext, it is enough to read through the subsequent stanzas (the context clues related to thought and emotion are highlighted in bold):

## II

What can I but enumerate old themes,  
First that sea-rider Oisín led by the nose  
Through three enchanted islands, allegorical dreams,  
Vain gaiety, vain battle, vain repose,  
Themes of **the embittered heart**, or so it seems,  
That might adorn old songs or courtly shows;  
But **what cared I** that set him on to ride,  
I, **starved for the bosom of** his fairy bride.

And then a counter-truth filled out its play,  
'The Countess Cathleen' was the name I gave it,  
She, pity-crazed, had given her soul away  
But masterful Heaven had intervened to save it.  
I **thought** my dear must her own soul destroy  
So did fanaticism and hate enslave it,  
And this brought forth a **dream** and soon enough  
This **dream** itself had all my **thought** and **love**.

And when the Fool and Blind Man stole the bread  
Cuchulain fought the ungovernable sea;  
**Heart** mysteries there, and yet when all is said  
It was the **dream** itself **enchanted me**:  
Character isolated by a deed  
To **engross the present and dominate memory**.  
Players and painted stage took all my **love**  
And not those things that they were emblems of.

### III

Those masterful **images** because complete  
Grew **in pure mind** but out of what began?  
A mound of refuse or the sweepings of a street,  
Old kettles, old bottles, and a broken can,  
Old iron, old bones, old rags, that raving slut  
Who keeps the till. Now that my ladder's gone  
I must lie down where all the ladders start  
In the foul rag and bone shop of the **heart**.

The context clues, highlighted in bold, are convincing proof that subtext does relate to the aspects of the text that are accessible to our intuition, and that it is not random or separate from the grammar string's substantiation in the language. In fact, all the four categories of the subtextual lexis point to the utmost significance of what has 'deserted' the elderly poet. And that is where we would leave it, had it not be for very substantial evidence that there is more justification to the first category ('significant people in one's life') than mere significance in general terms.

Let us look at the underlined text in the coda: 'Now that my ladder's gone...' This is a metaphor; moreover, 'ladder' obviously implies the same entity as 'circus animals' in stanza 1, because of the collocate 'gone'. Louw and Milojkovic (2014) provide an extensive corpus

analysis of the grammar string 'now that my \* is' in combination with 'gone'. The subtext reveals two sets of lexical variables: a) a significant person ('father', 'mother', 'son', 'husband', 'brother', 'wife', 'daughter', 'master'), and b) a quality essential to (honourable) existence ('mind', 'life', 'heart', 'time', 'hair', 'innocence', 'name', 'honour'). The 'ladder' is also relexicalised in the context of so many tangible objects located in the rag-and-bone shop. Regrettably, the analysis overlooks the existence of the other metaphor in stanza 1 which refers to 'circus animals'. Still, two metaphors in two very different sections of the poem (the opening stanza and the coda), employing different grammar strings when referring to the same thing, contain the same or very similar subtext. This proves that, unlike wordplay (at least in cases where the aim of the wordplay is simply to amuse, as in one-liners), metaphor consists of three, and not two, layers of meaning. Based on these examples, we can give a working definition of metaphor (which takes into account original metaphors, as in Yeats, unlike delexical metaphors discussed in Milojkovic (2016)):

Metaphor consists of a grammatical string containing a lexical item unique for the string in question (not found in the reference corpus).

The lexical item will interact with the surrounding context clues in two ways:

- a) relexicalisation will allow it to attain its literal meaning in the text and the reference corpus
- b) co-selection will allow it to attain a second, implied meaning in the text, which is the purpose of the metaphor; the states of affairs chunked in the reference corpus by co-selection will confirm this

The most frequent lexical items of the grammatical string in question (its subtext) will represent a third, hidden meaning. They will interact with the context clues in ways which will support the existence of the hidden meaning.

The two metaphors studied in this paper have something in common, besides the same (approximately) implied meaning and very similar subtext. The 'vehicle', or 'source' is, in both cases, a relexicalised entity of a much lower status than the implied. Circus entertains on a much more primitive level than poetry (which Yeats wrote, or aspired to write). The ladder, despite the

intertextuality (St John of the Ladder), is too sudden and, because of its collocates, too physical to be viewed respectfully – together with the poet who, we are told, used to be in need of one, and now has to do without. This depreciation, especially in the context of the title (the metaphor in the title is that of the circus animals, and not that of the ladder) is certainly tinged with self-irony – but given the extent of the defeat, the humour is bitter even if ‘ladders’ are co-selected with ‘start’ towards the end.

#### **4. Conclusion**

Corpus-derived subtext is a pervasive layer of meaning in texts, and its unravelling always contributes to the text’s interpretation. Still, a pun is a figure of speech which is sometimes intended to amuse. Therefore, grammatical auras embedded in it may not be relevant to the double meaning it conveys, if the double meaning is an end in itself, as is the case in comic one-liners. If, however, pun appears in serious texts, such as poetry, where both meanings may have their full realisations within the context of situation, subtext will be present and play its usual part, as a pervasive layer of meaning, within this text. For this reason, in the corpus-attested definition of pun, presented in this paper, there is no mention of subtext.

Within Yeats’s original metaphor, as it appears in the poem ‘The Circus Animals’ Desertion’, subtext plays a very significant role. Both metaphors studied in the article refer to the same implied meaning (which can be described as ‘literary devices’, in the broad sense). Although the grammar strings in both metaphors are very different (‘my \* were all’ vs. ‘now that my \* is’ co-selected with ‘gone’), the subtext in both is very similar. Not only is there a general aura of significance, but it contains, in both cases, a consistent group of the QPVs denoting a significant person, like a family member, for example. The consistency of subtext in the different grammar strings containing a lexical item with the same implied meaning, as shown in Yeats’s poem, confirms the ever-present role of collocation in meaning realisation, both in meaning creation and meaning construal.

## References

- Abrams, M. H. & Harpham, G. G. (2009) *A Glossary of Literary Terms*, Boston: Wadsworth Cengage Learning
- Chiaro, D. (1992) *The Language of Jokes: Analysing Verbal Play*, London: Routledge
- Collins Cobuild English Dictionary* (1998) London: HarperCollins
- Cudden, J. A. (1999) *The Penguin Dictionary of Literary Terms and Literary Theory*, London: Penguin Books
- Davies, M. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at <http://corpus.byu.edu/bnc/>
- Davies, M. (2008-) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>
- Davies, M. (2011-) *Google Books Corpus*. (Based on Google Books n-grams). Available online at <http://googlebooks.byu.edu/>
- Dickens, C. (1993) *The Pickwick Papers*, Ware: Wordsworth Editions Limited
- Goatley, A. (2013). *Meaning and Humour: Key Topics in Semantics and Pragmatics*, Cambridge: CUP.
- Hunston, S. (2007) 'Semantic Prosody Revisited', *International Journal of Corpus Linguistics* Vol. 12, No. 2, pp. 245-264.
- LaPolla, R. J. (2006). 'The How and Why of Syntactic Relations', invited plenary address and keynote of the Centre for Research on Language change Workshop on Grammatical Change at the Annual Conference of the Australian Linguistics Society, University of Queensland.
- Lodge, D. (1984) *Small World*, Harmondsworth: Penguin Books.

Louw, W. E. (1993) 'Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies', in M. Baker et al. (eds.) *Text and Technology: In Honour of John Sinclair*, Amsterdam: John Benjamins, pp. 157-176.

Louw, W. E. (2008) 'Consolidating empirical method in data-assisted stylistics: towards a corpus-attested glossary of literary terms', in S. Zyngier, M. Bortolussi, A. Chesnokova and J. Auracher (eds.) *Directions in Empirical Literary Studies. In Honour of Willie van Peer*, Amsterdam: John Benjamins, pp. 243–264.

Louw, W. E. (2010a) 'Collocation as instrumentation for meaning: a scientific fact', in W. van Peer, V. Viana, and S. Zyngier (eds.) *Literary education and digital learning: methods and technologies for humanities studies*, Hershey, PA: IGI Global, pp. 79-101.

Louw, W.E. (2010b) 'Automating the extraction of literary worlds and their subtexts from the poetry of William Butler Yeats', in M. Falces Sierra et al (eds.) *Para por y Sobre Luis Quereda*, Granada: Granada University Press.

Louw W. E. (2014) 'The analysis and creation of humour', in Coloru, O. & Minunno, J. (eds.) *L'Umore in prospettiva interculturale: Immagini, Aspetti e Linguaggi/ Crosscultural Humour: Images, Aspects and Languages*. Parma, Atelier 65. (Paper 10 in Appendix).

Louw, W. E. & Milojkovic, M. (2014) 'Semantic Prosody', in P. Stockwell & S. Whiteley (eds.) *The Cambridge Handbook of Stylistics*, Cambridge: Cambridge University Press, 263-280.

Louw W. E. & Milojkovic, M. (2016) *Corpus Stylistics as Contextual Prosodic Theory and subtext*. Amsterdam: John Benjamins

Milojkovic, M. (2011) 'Semantic prosody and subtext as universal, collocation-based instrumentation for meaning and literary worlds', in V.P. Zakharov (et al.), *Труды международной конференции «Корпусная лингвистика – 2011»*. Санкт-Петербург: С.-Петербургский гос. университет, Фил. Факультет, pp. 47-52

Milojkovic M. (2016) A Corpus-Attested View of Business English Metaphors. In Moreno Ortiz A. and Pérez-Hernández C. (eds), *CILC2016. 8th International Conference on Corpus Linguistics*. EPiC Series in Language and Linguistics, volume 1. 283-297

Milojkovic M. & Louw W. E (2017) 'Towards a corpus-attested definition of creativity as accessed through a subtextual analysis of student writing. In R. Breeze and C. Sancho Guinda (eds.) *Essential Competencies for English-medium University Teaching*, Springer, pp. 125-141.

Simpson, P. (2003) *On the Discourse of Satire: Towards a Stylistic Model of Satirical Humour*, Amsterdam: John Benjamins.

Sinclair, J. (1987) *Looking Up: Account of the Cobuild Project in Lexical Computing*, London: Collins Cobuild

Xiao, R. & McEnery, T. (2006) 'Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective', *Applied Linguistics*, Vol. 27, No. 1, pp. 103-129.