ISSN 2299-5900

TOKEN

A Journal of English Linguistics

Volume 1/2012

Token: A Journal of English Linguistics
Volume 1

This volume is dedicated to
Professor Dieter Kastovsky (1940-2012)

JAN KOCHANOWSKI UNIVERSITY OF KIELCE

Token: A Journal of English Linguistics

Volume 1

Edited by

Sylwester Łodej
John G. Newman

Jan Kochanowski University Press

Kielce 2012

Editorial Correspondence
Uniwersytet Jana Kochanowskiego w Kielcach
Instytut Filologii Obcych
25-310 Kielce
ul. Tadeusza Kościuszki 13
Poland
www.ujk.edu.pl/token
token@ujk.edu.pl

Token is the original medium of publication for all articles that the journal prints.
Token is not an electronic publication.
This publication has been supported financially by the Polish Ministry of Science and Higher Education.

Parasitic passives of intransitives in English

John M. Anderson
University of Edinburgh (Emeritus)

ABSTRACT

There is an unusual construction in English involving the passive of a normally agentive intransitive verb, such as *attempt*. The viability of such a passive depends on its being complemented by a normal passive of a transitive. The intransitive passive can be said to be parasitic upon the transitive. Thus, the overall construction takes the form of *They were attempted to be dismissed*. The present essay, starting from a couple of attested examples, explores the structure of such sentences and attempts to account for the parasitic requirement that ensures the acceptability of the intransitive passive concerned. Within a framework based on recent developments in notional grammar, it offers analyses of passives of various types in English as a background to establishing an understanding of the nature of such parasitic passives.

1. Introduction

The examples in (1) illustrate the unusual construction I am concerned with here:

(1) a. … we have often relieved them, both with Victuals and Cloaths too, even while they were pretended to be kept by their barbarous Aunt. (from *Roxana* by Daniel Defoe, p.186 in the Folio Society edition)

b. … most of their carriages were so decayed that that they could not be attempted to be fired. (from *The Ionian Mission* by Patrick O'Brian, p.161 in the Folio Society edition)

The subordinate clauses in these both exhibit the passive of an intransitive that takes a passive transitive as complement; the complement cannot be active. The subject of the passive intransitive is semantically a complement of the subordinate transitive – specifically the complement traditionally labelled as 'object'. The intransitive of course has no 'object' of its own, as illustrated in (2), where the verb is complemented by a subject and an infinitive phrase:

(2) a. Their aunt pretended to keep them.

b. They could not attempt to fire them.

The complement that is 'raised' in (1) to occupy the subject position in the relevant passive clauses is 'borrowed' from the subordinate transitive. It is for this reason that I refer to these examples as involving a 'parasitic passive'.

Even when such verbs have transitive argument structure, with an 'object' nominal rather than an infinitive, as in (3), their 'objects' are not normally prototypical, discrete, concrete nominals, but ones associated semantically with 'verbal' notions such as 'action' or 'mental state':

(3) a. Bobby pretended sleep/indifference.

b. Frodo attempted a dive/persuasion/prevarication.

Indeed, as illustrated here, in many cases the 'object' is patently deverbal. And constructions involving both verbs often rely on a lexical periphrasis to spell out the verbality, as in (4):

(4) a. Bobby pretended to have a sleep.

b. Frodo attempted to do a dive.

To that extent the verbs are already parasitic in taking the 'objects' in (3) and (4); as transitives they depend on the presence of 'verbality' in their 'objects'. They are reluctant transitives. But the expressions in (1) illustrate a more serious incidence of parasitism in achieving transitivity.

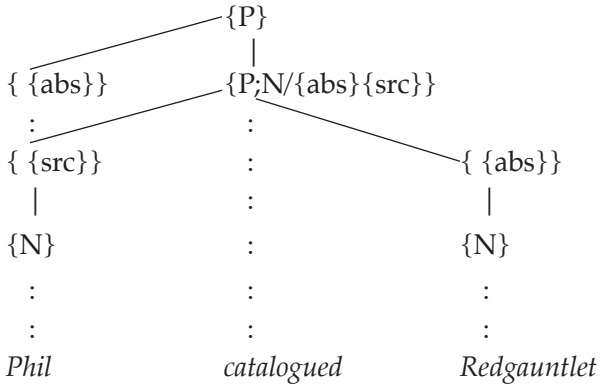
Let us look more carefully at the structure of the sentences in (1), in the context of the account of 'raising' and passive offered in recent presentations of notional grammar (particularly in Anderson 1992: §4.1, 1997: 202-6, 248-9, 2006: §§9.2.2, 12.2.2, 2011: I, §5.4; II, §§3.3.4, 3.5; III, §7.1, Böhm 1993: §§3-4).

The analysis of passive that is appropriate within the assumptions of this context is spelled out in the section that now follows.

2. Canonical passives in English

The canonical passive verb is a form of a verb that takes an argument that is the **source** of the event or state being represented – roughly, an 'agentive' or an 'experiencer', as these latter terms have been used of late – and a distinct **absolutive**, or 'neutral', argument. In the active the former is subject and the latter 'object'. This is illustrated as part of the dependency structure in (5):

(5)



The syntactic tree in (5) is a dependency tree with root {P}, the finiteness primary category. {P;N} is the categorization for lexical verbs, in which the predicative feature **P** governs the argumental feature **N**, as indicated by the semi-colon. In (5) this verb is shown as having been **converted** to a finite in the lexicon, to which it is joined by a dependency arc that is not associated with linear difference between the head and the dependent; the latter is said to be **subjoined** to the former. Similarly, the two {N} categories, introducing arguments, have each been converted, lexically subjoined, to a **functor**. Functor is the primary category that is unspecified but bears semantic relations as secondary categories, in this case source ({src}) – agentives in this instance – and absolutive ({abs}). Functors can appear independently as prepositions in English, just as the finiteness category {P} can appear as functional verb forms like *may* or *is*.

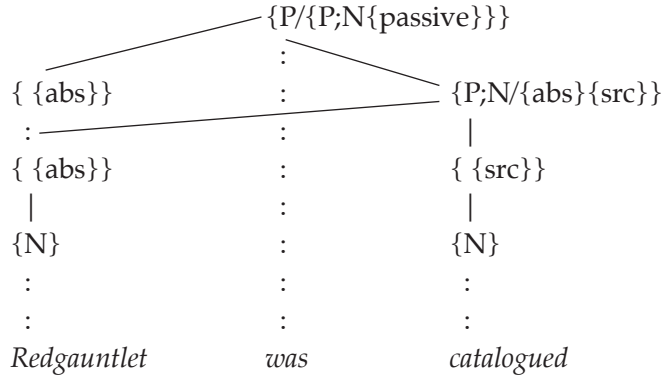
The non-vertical solid lines represent dependency arcs introduced in the syntax in building the syntactic structure that satisfies the demands of the individual words. In this case the head and the dependent occupy

distinct positions; the dependent is **adjoined**. The source and the absolutive satisfy the valency of the lexical verb, shown to the right of the slash, '/', in its representation; this satisfying permits the introduction of the syntactic dependency arcs. The absolutive that bears a dependency relation to the {P} in (5) is not specified as part of the valency of {P}, however; it is a **free absolutive**. It is present in response to a requirement that any verb, functional or lexical, must have a syntactically dependent absolutive, whether or not this is part of its valency. As a functor, this absolutive needs a dependent; this is true of all functional categories, such as {P}, {N}, and { } (functor). To satisfy this requirement, the free absolutive in this instance shares the argument of the source, which is 'raised' to be associated with it; this is shown by the (discontinuous) association line linking them. The free absolutive shares that argument of the {P;N} whose semantic relation is highest on the **subject-selection hierarchy**; this is the designated subject. The present case illustrates that the source outranks the absolutive. This argument thereby comes to occupy a position in front of its verb, which is that assigned to free absolutives, whereas normally in English the arguments of a verb come after. The position of the shared argument is determined by the norm for the uppermost functor, here the free absolutive of {P}, which precedes its head.

I have said that {N}, determiner, in common with other functional categories, takes a complement, which may be either lexical or syntactic. Typical complements are nouns (*the book*), and names and pronouns, with in English the latter being converted to determiners in the lexicon. I have, however, left the structure of the {N}-phrases unexpressed in (5), since this is not relevant to our concerns here. Clearly much else is omitted, for similar reasons, from the representation in (5). For a fuller picture I refer the reader again to the works cited in the introduction.

But let us now turn to the representation of the passive 'equivalent' of (5). Firstly I give the structure for the 'short passive', as in (6):

(6)



Here there is an unspecified source argument that has been lexically subjoined to {P;N}, in satisfaction of that part of the valency of the latter. It has been lexically **incorporated**, as part of the formation of this form of the verb. As a result, the absolutive argument of *catalogued* is highest on the subject-selection hierarchy, and thus is shared with the free absolutive of the independent {P}. The latter lacks an absolutive from its valency, as with the {P} in (5).

Both subject-formation in (5) and passivization in (6) can be said to involve so-called 'subject-raising', i.e., in present terms, sharing with the free absolutive of {P}. The *catalogued* in (6) is a form of the verb that cannot be converted lexically to a finite. To achieve finiteness it serves as the complement of an independent functional finite, *was*, that is dedicated to the introduction of passives. 'Passive' in (6) is a simplifying shorthand for the form of verb that, associated with the incorporation of the 'usual' subject, promotes to subject status the second-highest argument of the verb.

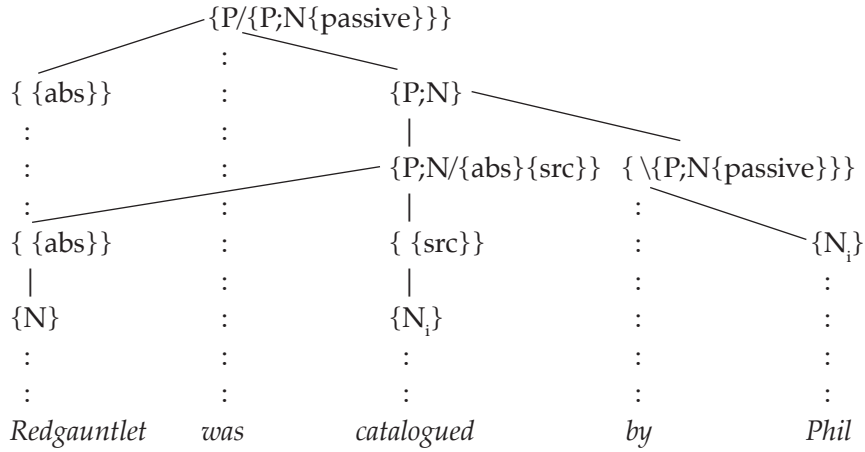
Not all lexical incorporations of the titular subject of a verb are associated with transfer of subject status away from it. Compare e.g. (7), with incorporation of the subject of *keep*, which shows no sign of 'displacement' as such:

(7) Mother says to keep it.

The 'transfer' of subjecthood is a property of the passive form. The construction in (7) will be of interest subsequently, however.

In the 'long' passive in (8) the lexical items in (5) recur, but not quite with simply the same semantic relations:

(8)



By Phil is a functor phrase dedicated to modifying a passive verb: the **modification** relation is indicated by backward slash, '\', and it has the effect

of introducing a replica of the modified category above that category. This creates a more inclusive construction containing the modifier. The duplicate {P;N} is transparent to the valency requirement of {P}. The functor heading the modifying phrase takes as a complement a {N} – in common with other functors – but this {N} is marked as co-referential with the incorporated {N} of *catalogued*. The modifier provides an optional further specification of the incorporated source argument.

3. Passive plus infinitive

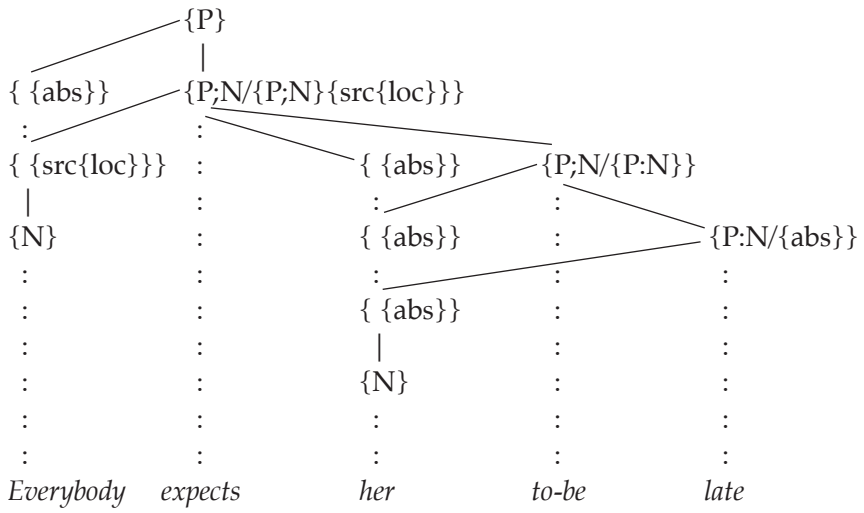
The passives in (1) involve dependent infinitive constructions. I look now at how the canonical passive interacts with the infinitive. A typical infinitive-taking verb is the familiar *expect* of (9), respectively in the passive and active forms:

(9) a. She is expected (by everybody) to be late.

b. Everybody expects her to be late.

The active in (9b) can be represented as in (10):

(10)

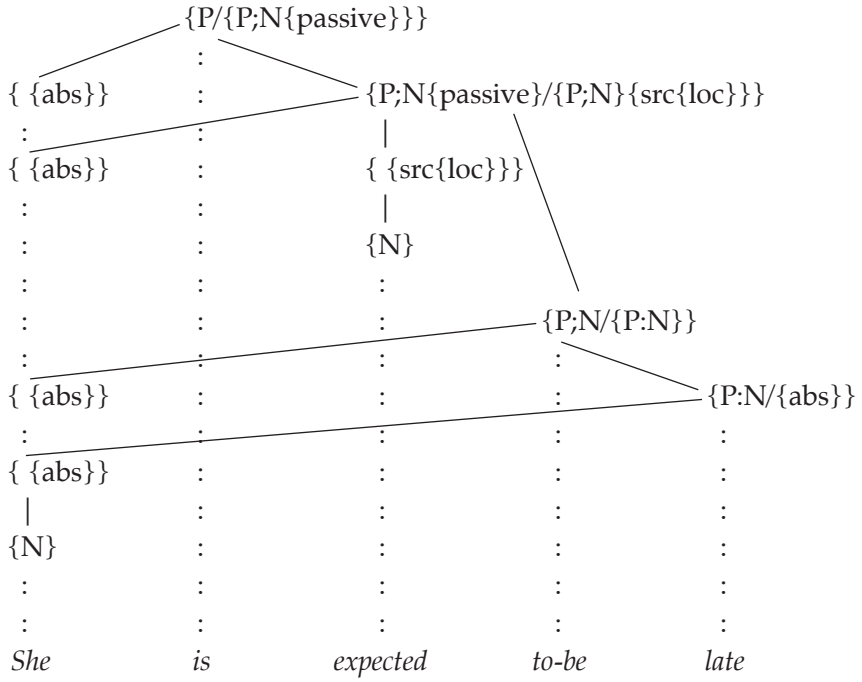


{P:N} is an adjective – here *late*. The colon in the representation for an adjective signifies that the two features, **P** and **N** are equipollent; nouns are {N;P}. The adjective in (10) is the complement of the infinitive *to-be* (whose internal structure is not relevant here).

The adjective has absolutive in its valency; but the other two absolutives immediately preceding *to-be* and realized as *her* are free absolutives. *Her* is the subject of *late*, so, as in (6), it undergoes 'raising' to share its argument with the free absolutive of the governing predicator, {P;N/{P:N}}. And this in turn shares with the free absolutive of *expects*. This creates the pattern of associations in (10) above *her*. Again the highest sharing functor determines the position of the shared argument. In this case the highest functor is the free absolutive of the {P;N} *expects*, which, as a {P;N} rather than a {P}, takes its free absolutive to the right, as is usual with complements of a verb in English. We have so-called 'object-raising'. The only other semantic relation in (10) that appears in a valency is the { {src{loc}}}, a secondary source plus a tertiary locative, which is the specific representation for an 'experiencer': this characterizes the subject of *expects*, which also shows sharing, with the free absolutive of the root {P}.

Even though the absolutive dependent on the {P;N} of *expects*, realized as *her*, is not part of its valency, it is eligible for being the subject of a passive with *expect* by virtue of the dependency relation between the upper {P;N} (*expects*) and its free absolutive. The verb is derivatively transitive. Thus, we find the viable passive in (9a), the structure of whose 'short' version we can represent as in (11):

(11)



The pile-up of free absolutives on the left reflects the subsuming of subject-formation, passivization and more traditional 'raising' as the uniform sharing with a free absolutive of {P} of the argument designated by subject-selection.

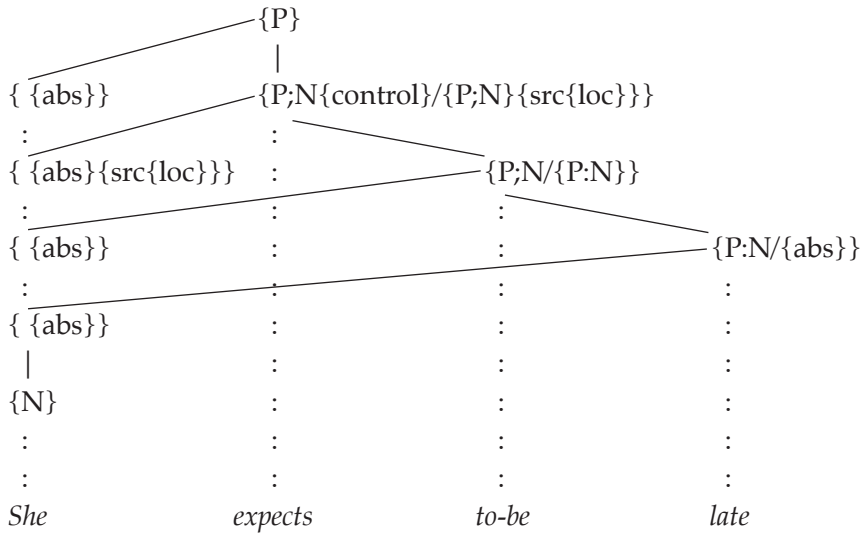
Infinitive-governing *expect* is not always a 'raising' verb, as is illustrated by (12), which exhibits instead 'control':

(12) She expects to be late.

Here too I shall follow the kind of analysis offered in the works invoked in the Introduction, whereby obligatory 'control' is another instance of argument-sharing with a free absolutive. But the 'control' option is the marked possibility for introduction of the free absolutive, and is appropriately singled out by presence of the semantic feature {control}. The syntactic effect of {control} is to prevent the free absolutive from being introduced independently, but instead in combination with an argument required by the valency of the verb.

In those terms we can associate with (12) the representation in (13)

(13)



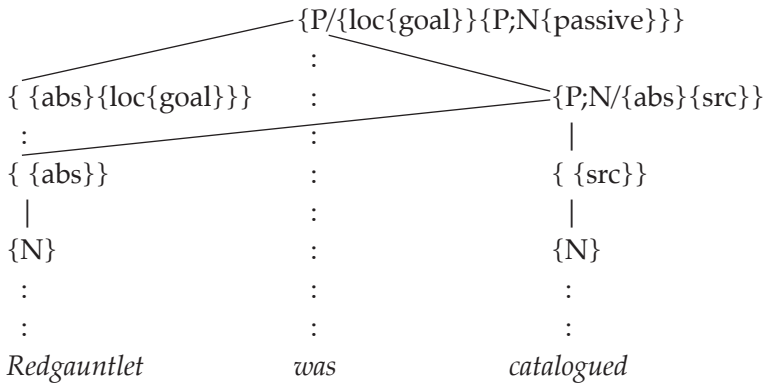
Here the free absolutive of *expect* is not introduced as an independent functor, but is associated with the semantic relation that does appear in the valency of *expect*, the 'experiencer' we also find in (9), again represented as '{ {src{loc}}}'. But it retains its capacity, as a free absolutive, to host subjects from subordinate {P;N}s, as expressed in (13). 'Control' also involves

argument sharing. The introduction of 'control' leads us on to a so-far-neglected aspect of passive.

4. Passive and 'control'

The passive structure I have so far assumed is somewhat simplified. Anderson (2006: §12.2.2) argues that passive is also a 'control' construction, analogous to that I have just looked at. A still slightly simplified version of the view described there is embodied in the representation of passive offered in (14).

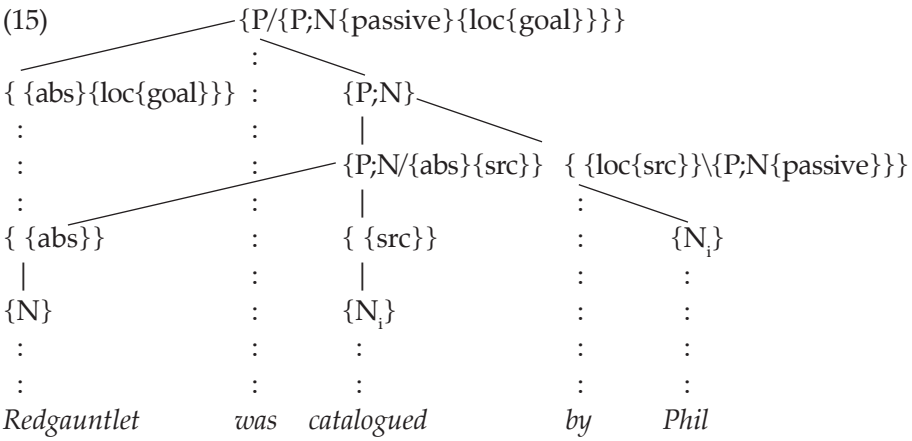
(14)



Here passive *be* is treated as a 'control' verbal whose valency contains a goal locative (a 'receiver' – a directional specialization of 'experiencer'), and the free absolutive is associated with it. I assume that the passive auxiliary is redundantly a control word, so I have left out {control} in the representation in (14).

Such a proposal has a long history. Anderson (1972: §2), in defending a precursor of the kind of analysis embodied in (14), points to earlier work illustrating the directionality of passive constructions. In such constructions in other languages – and in the English *get*-passive – the 'recipient' character of the passive verbal is more transparent in the form of the verb corresponding to *be* in (14). And compare, on the semantic character of the passive, Gildersleeve and Lodge's (1867/1968: §112.2) succinct formulation, for example: '[t]he Passive Voice denotes that the subject receives the action of the verb'.

Similarly, the optional modifier with passive is arguably a locative source, as embodied in (15):



As Anderson (1972: §2) again points out, in various languages this modifier is headed by a functor that is otherwise a transparent locative source (possibly marked by an inflectional ablative); in others the form is historically an actional source ('agentive') or 'instrumental' or 'path' – the latter two combining goal and source. In (15) the modifier is interpreted as a locative source – though this is not crucial to our present concerns.

So much for the canonical passive of English. It is not my purpose here to investigate the full variety of structures that might be described as 'non--canonical passives' in English, including, for instance, a range of locative constructions. Still less shall I attempt to deal with the range of constructions that might be described as 'passive' cross-linguistically (some of which is documented by Postal (1986), for example). My stated concern is with the passive construction illustrated in (1). But I shall approach this via a somewhat less divergent variety of non-canonical passive, as will be described in the immediately succeeding section. Part of the interest of this variety of passive is the need not to confuse it with the more specifically parasitic passives in (1).

5. Dedicated intransitive passives

What I have in mind here are passives such as that in (16a):

(16) a. He is said (by everybody) to be depraved.

b. *Everybody says him to be depraved.

As is well-known, the active 'equivalent' of such sentences, illustrated in this case by (16b) is unacceptable. At first sight, it looks as if what should be

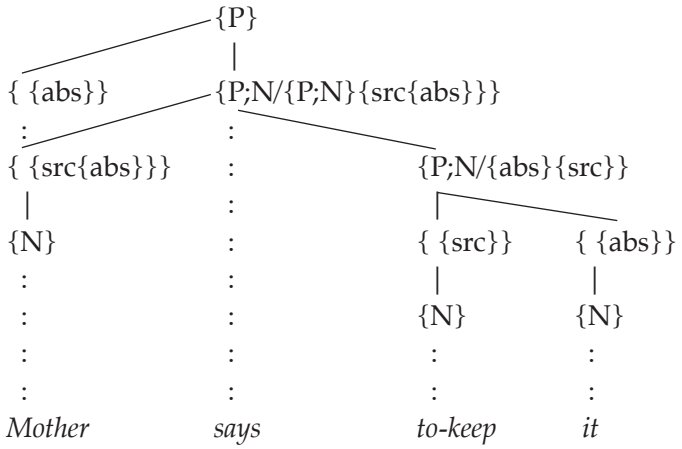
our concern here is what has 'gone wrong' with (16b). But closer inspection suggests otherwise.

We have a rather different situation from that illustrated by the infinitive constructions in §3 in the case of the *say* of (16) – and of (7):

(7) Mother says to keep it.

(16b) gives no evidence, in the form of raising, of the presence of a free absolutive with the active form of that verb. And nor does (7), which is not a 'control' structure; the lower verb has an incorporated independent subject. Thus, the *say* verb of (7) is intransitive: the subject of *say* is its only non-verbal argument, and it is '{ {src{abs}}}' in terms of the present notation. So that we might represent (7) as in (17):

(17)



Say seems to be an agentive intransitive that takes a verbal complement. There is, of course, a related verb *say* that takes 'cognate objects', as exemplified in (18a):

(18) a. She will say a few words/nothing/whatever he suggests.

b. She will sing a few stanzas/nothing/whatever he suggests.

Compare the similar behaviour of the basically agentive intransitive *sing* of (18b). Such instances of *say* as (18a) do not seem to be relevant to the phenomena under discussion, except that they, being transitive, have a passive 'equivalent'. *Say* is ambivalent in transitivity, depending on whether it takes an infinitive or not.

However, (16a), is passive of an intransitive verb, unusual in English. There are other languages, such as Dutch, which in general show passives of agentive intransitives, as illustrated by (19):

(19) Er wordt hier door de jongelui vaak gedanst.
[it was here by the young.people often danced]

(Perlmutter 1978). But in this case the free absolutive is filled by an expletive, given that the subject has been absorbed. Such an expletive is lacking in (16a). There is indeed a passive construction of intransitives involving such verbs as *say* that does show an expletive, but again, as in (16a), with only a non-nominal complement, other than the subject. This is exemplified by (20a), with, unlike (16a), a full clausal complement:

(20) a. It is said (that) he is depraved.

b. They say (that) he is depraved.

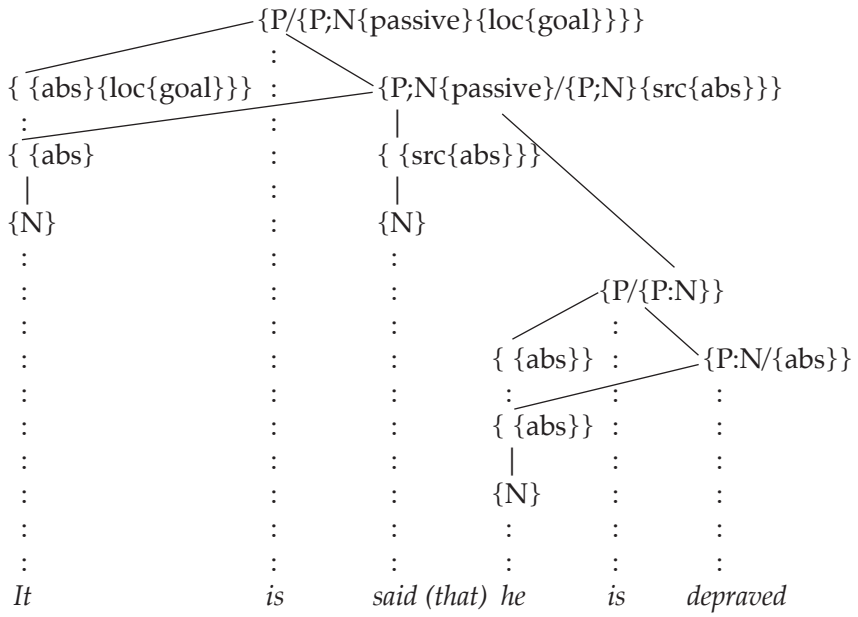
c. *(That) he is depraved is said.

The main verb in (20a-b) is not transitive, any more than that in (19). The subordinate clause is not an 'object' so that it doesn't appear in subject position in the passive construction in (20c). Only the expletive *it* in (20a) is available as subject.

Say, exceptionally for an agentive intransitive, has a passive form in (21), perhaps under the influence of the related cognate-object verb of (18a). The main verb here shares the agency property with the corresponding argument of the prototypical (transitive) passive in (6). But the *said* verb is deprived of this agentive, its only independent nominal complement, by passive-formation. The valency is satisfied by a lexically-incorporated intransitive agent; and the passive form of the verb is supplied with a free absolutive. There is no complement of the *say* verb available for subject-formation, and the passive subject of *depraved* is accommodated internally to the lower clause, by the lower {P}. The free absolutive of *said* is hosted by the passive goal free-absolutive of the upper clause, but its own valency is unsatisfied. And the whole configuration on the left in (21) is realized by the expletive {N} *it*.

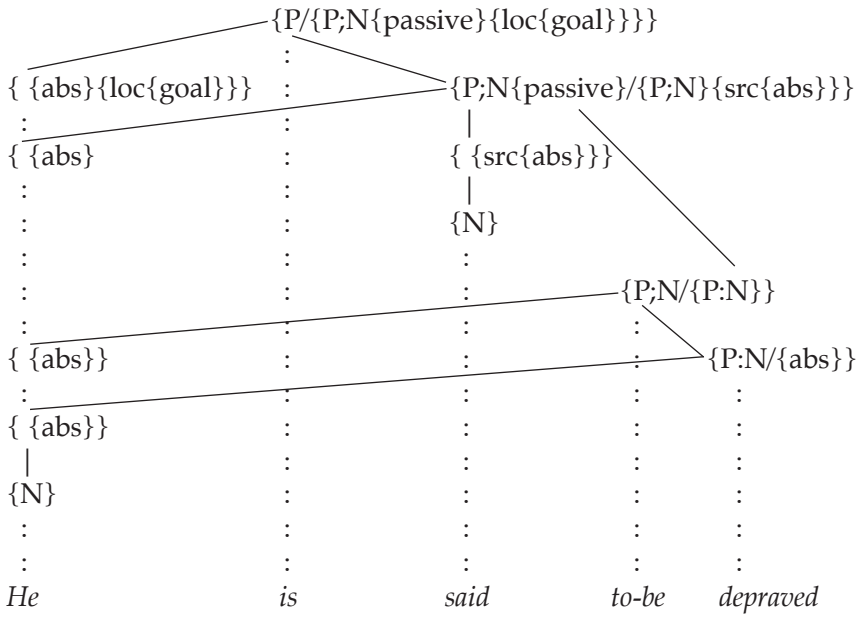
We can thus represent the structure of (20a) as in (21), where I ignore the status of the optional *that*:

(21)



But there is no expletive in the case of the construction with infinitive complement in (16a). Instead, the free absolutive of the passive verb can be satisfied by the subject of the infinitive, unlike in (21) where the subject of the lower verb is hosted by the free absolutive of the lower {P}, and is not available for raising. Thus we can represent (16a) as in (22), where, unlike in (7) or (21), we have raising:

(22)



The passive form again involves incorporation of the argument highest on the subject-selection hierarchy. And again this leaves the *said* form without overt argument. This is associated with the introduction of a free absolutive that in this instance is satisfied by the subject of the subordinate *to be depraved*. This construction thus seems to be a passive form apparently dedicated to offering initial – i.e. unmarked topical – position to the subject of a dependent verb.

We set out in this section with wondering what has 'gone wrong' with (16b), given the acceptability of (16a). But the non-acceptability of (16b) is just what we would expect, given that *say* is an intransitive agentive, which normally do not passivize in English. And it is (16a) that is exceptional, in terms of *say* having a passive form. This is not extended to agentive intransitives in general in English, but limited to verbs with appropriate semantics. What we turn to now is a construction that, as a dedicated passive, resembles that in (16a); but it is an even more restricted variety of passive, again involving intransitives, but requiring the presence of a canonical passive as complement. This brings us back to the construction illustrated in (1), containing what I called 'parasitic passives'.

6. Parasitic passives

The constructions that were cited by (1) are similar in showing the passive of an agentive intransitive, but, further, the presence of this seems to be associated with the presence of a dependent passive construction:

(1) a. … we have often relieved them, both with Victuals and Cloaths too, even while they were pretended to be kept by their barbarous Aunt.

b. … most of their carriages were so decayed that that they could not be attempted to be fired.

Pretend, unlike *say*, is normally a 'control' verb, as in either sentence in (23):

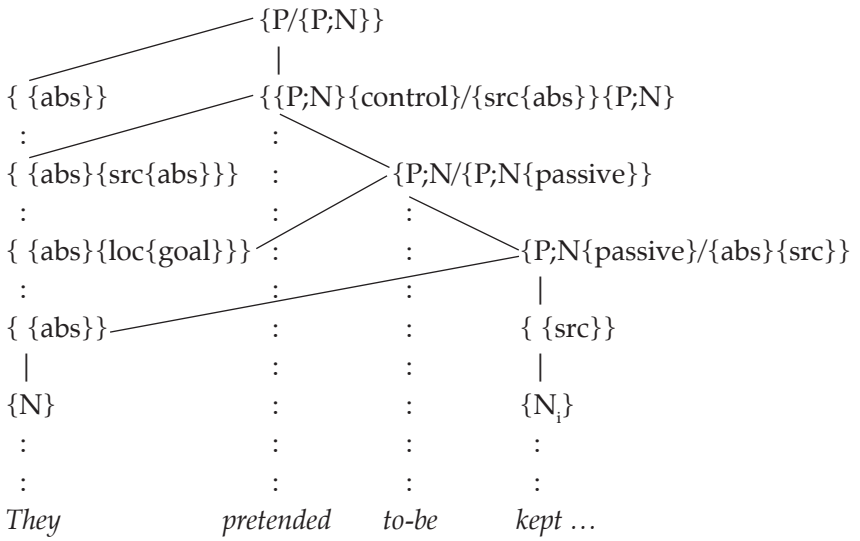
(23) a. Their barbarous aunt pretended to keep them.

b. They pretended to be kept (by their barbarous aunt).

Attempt is also such a 'control' verb. But I shall, for simplicity, illustrate what follows with reference to (1a).

We can represent the short version of the sentence in (23b), for instance, as in (24):

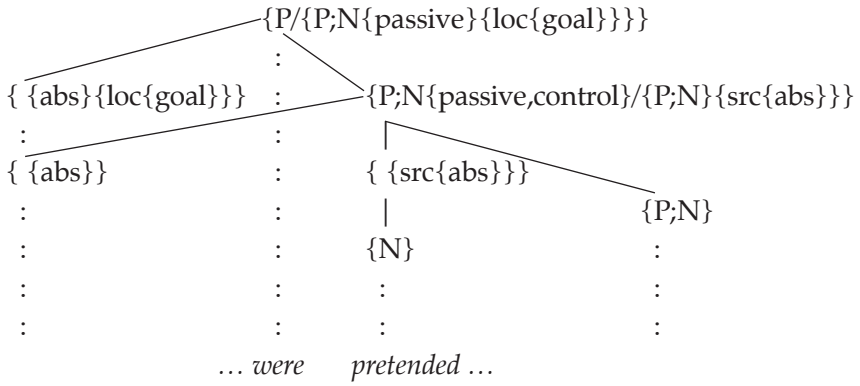
(24)



The agentive 'controller' of *pretended* hosts the passive 'controller' of *to-be*, which in turn hosts the absolutive of the passive verb *kept*. In (23a) the agentive 'controller' hosts the agentive subject of *keep* directly.

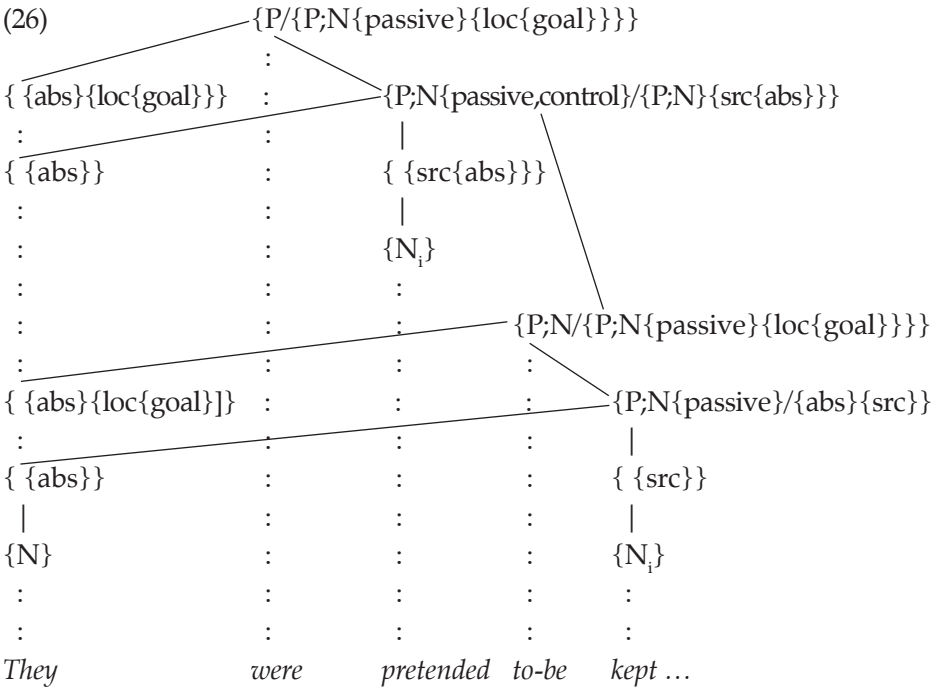
The passive form of *pretend* in the *while*-clause in (1b), however, removes from the valency of that verb the need for an independent 'controller', since such has been incorporated into the verb; and there is provided instead an independent free absolutive which is hosted by the free absolutive of the passive {P}, as shown in (25):

(25)



Normally, the 'controller' would share with the subordinate subject, as in (23). But we now have an incorporated 'controller', and apparently a 'raising' construction, with a free absolutive of the *pretended* verb looking for a subject to host. But, semantically, as indicated by the presence of {control}, the incorporated 'controller' is also still looking for a 'controllee'. We have an apparent conflict. There is an incorporated 'controller' in search of a 'controllee', and a free absolutive needing a lower subject to satisfy it. This can be resolved if, as in (25), the clause subordinate to *pretend* is also passive. The subject of the lower passive satisfies the free absolutive of the *pretend* clause and 'control' is expressed by marking of the incorporated controller as co-referential with the incorporated argument of the lower passive verb. Only a passive complement can provide for the satisfaction of the demands of the parasitic passive.

The appropriate structure for (1b) is presented in (26), where I have, however, omitted the final modifying phrase, as not pertinent:



In this case all of the arguments that aren't predicators are on the left, realized as *they*. The lowest absolutive, satisfying the absolutive valency of *kept*, is linked by a couple of free absolutives to the highest free absolutive, that of the parasitic passive; and this is possible only because of the passivization of

kept. And the co-reference of the 'pretender' and the 'keeper' is expressed in the incorporated arguments of these passive verbs, thus completing the satisfaction of the demands of the parasitic passive.

A final, obvious observation is that since the construction exemplified in (1) is recursive, as illustrated by (27), the parasitism is in principle indefinitely extensible:

(27) They were pretended to be attempted to be fired.

In the cases of both (1) and (27) the legitimacy of the passive of the agentive intransitives depends on the (ultimate) presence of a normal subordinate passive; their legitimacy is parasitic. And again, a major motivation for this overall construction appears to be the fronting of an argument, in this instance the absolutive argument of the lowest verb. However, the parasitic passive construction, both in terms of its manifestation in Present-day English and as concerns its historical evolution, merits much more extensive study.

REFERENCES

Sources

Defoe, Daniel
1724 *Roxana: The Fortunate Mistress*. London: T. Warner. [Repr. 2010] London: The Folio Society.

O'Brian, Patrick
1981 *The Ionian Mission*. London: William Collins. [Repr. 2010] London: The Folio Society.

Special studies

Anderson, John M.
1972 "The hierarchy of quasi-predications", *Revue roumaine de linguistique* 17, 23-44, 121-40, 193-202, 319-35.
1992 *Linguistic Representation: Structural Analogy and Stratification*. Berlin: Mouton de Gruyter.
1997 *A Notional Theory of Syntactic Categories*. Cambridge: Cambridge University Press.
2006 *Modern Grammars of Case: A Retrospective*. Oxford: Oxford University Press.
2011 *The Substance of Language*. 3 vols. Oxford: Oxford University Press.

Böhm, Roger
1993 "Predicate-argument structure, relational typology and (anti)passives: towards an integrated case grammar account". Duisburg: L.A.U.D., Series A, No.336.
Gildersleeve, B.L. – Gonzalez Lodge
1968 *Gildersleeve's Latin Grammar* (3rd edn.). London: Macmillan. Perlmutter, David M.
1978 "Impersonal passives and the unaccusative hypothesis", *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* 4, 157-89.
Postal, Paul M.
1986 *Studies of Passive Clauses.* Albany NY: State University of New York.

Sense and sensibility: Verbal morpho-syntax in nineteenth-century Scottish emigrants' letters and the intersection of standard and vernacular usage

Marina Dossena

University of Bergamo

ABSTRACT

This paper discusses verbal morpho-syntactic features in a corpus of nineteenth-century letters written by encoders of Scottish origin, whose levels of education vary quite considerably. The aim is to identify the features which appear to be most resistant to anglicization, paying particular attention to instances of the so-called Northern Subject Rule, and to modal auxiliaries. The aim is to shed more light on Late Modern English beyond the materials currently available for the study of this variety, especially as far as Scotland is concerned. Data compiled to date will be examined against data available in the recently-launched *Corpus of Modern Scottish Writing, 1700-1945*.

1. Introduction

While written documents typically reflect an attempt to imitate standard models of educated discourse, letters place themselves at the intersection of formal modes of address and much less formal ways of conveying personalized and intimate contents. As observed in other studies (e.g., Dossena 2008a and forthcoming a), it is therefore possible to use such documents as witnesses of vernacular usage, where less monitored linguistic choices derive from the greater importance given to the immediacy of the message and to the often powerfully emotional relationship existing between encoders and recipients. In particular, it may be interesting to focus on texts written by minimally-educated encoders; predictably, they are not written in dialect, but they do include regional features, whether in spelling or syntax, which may help the investigation of socio-geographical variation.

The decision to concentrate here on verbal syntax and morphology in popular writing is meant to shed more light on Late Modern English beyond the materials currently available for the study of this variety, especially as far as Scotland is concerned. My analysis will rely on and compare materials of the *Corpus of Nineteenth-century Scottish Correspondence* (19CSC, currently in preparation at the University of Bergamo: see Dossena 2004 and 2006 and Dossena – Dury 2008) and the nineteenth-century "personal writing" section of the *Corpus of Modern Scottish Writing* (CMSW), comprising ca. 513,000 words and accounting for ca. 9% of the total.

The *Corpus of Nineteenth-century Scottish Correspondence* is a collection of letters written by (or on behalf of) men and women of varying ages, from different social backgrounds, and for different purposes. It aims to be a "second-generation corpus" (Mäkinen 2006): i.e., one allowing clearly defined, focused studies in which scholars can concentrate on relatively few authentic texts, rather than edited ones, in order to highlight the specificities of linguistic traits without the risk of interfering "noise" created by editorial choices (Lass 2004). Indeed, as shown by Kohnen (2007 and 2008), qualitative analyses of a homogeneous corpus may show a validity that goes beyond the statistical one, especially when the object of investigation is barely quantifiable, on account of its elusiveness or intrinsic variability (such as that of speech acts, evaluation and modality).

At the moment of writing (October 2011) 19CSC comprises ca. 450 letters (between drafts, fair copies[1] and archival copies, equally distributed between familiar correspondence and business letters), for a total of ca. 120,000 orthographic units. The structure of the corpus, therefore, addresses

1 By "fair copies" we mean letters that were actually sent. In some cases, fair copies were in fact the first and only draft, as shown by the self-corrections appearing in them, as time or financial constraints did not afford encoders preliminary drafts.

I gratefully acknowledge permission to quote from MSS held in the Glasgow University Archives, the National Archives, the National Library of Scotland, and the Archives of the Bank of Scotland in Edinburgh, and the Thomas Fisher Rare Book Library in Toronto. Such permission does not extend to third parties, so the quotations presented in this paper should not be used elsewhere. I am also indebted to Richard Dury for his help in the design and compilation of 19CSC, and for valuable comments throughout the investigation process. In the examples most names, locations and dates are omitted for reasons of privacy. Line and page break indicators (# and ### respectively) are normally omitted for reasons of space and legibility.

A preliminary version of this paper was presented at the 14th Methods in Dialectology Conference (London, ON, 2-6 August 2011), within the workshop on "English and European Historical Dialectology" organized by the author.

a few interesting methodological questions, some of which augment those addressed by Jucker – Taavitsainen (2008). From the terminological point of view, for instance, the label "orthographic unit" is deemed to be preferable to "word", because in eighteenth- and nineteenth-century manuscripts words often coalesced orthographically, and introducing artificial breaks would disrupt the authenticity of the transcription (see Lass 2004 and Dury 2006 and 2008). Also, I prefer to use the term "encoder", instead of "writer", as the latter can only apply to holograph letters. When the contribution of an amanuensis cannot be excluded, or is in fact expected, as in the case of managers dictating to secretaries, the person who actually "writes" the letter is not necessarily the person whose meanings are conveyed (see Dossena, forthcoming b). As for "sender", this may also be appropriate, though of course messages could additionally be conveyed on behalf of subjects other than the one who actually sent the letter. The term "recipients" is considered preferable to that of "addressees", as the person to whom the letter was actually addressed was not necessarily the only person who read it – in fact, as is well-known, in Late Modern times letters were often circulated among friends and family. This is especially true of emigrants' letters, as these often included messages to / from participants other than the individual encoder and specific addressee.

This corpus of correspondence is further supplemented with transcriptions of diaries written by people whose linguistic competence is comparable with that of the letter writers. The aim is to have a relatively wide range of encoders, from the least schooled to the best educated, both men and women, of various ages and differing social classes, so that phraseological variation, syntax and pragmatic features may be studied across registers and styles: not only "standardized" ones, but also those that may approximate spontaneous (and perhaps less educated) usage.

The modular structure of 19CSC has already enabled a series of studies on different aspects of Late Modern Scottish correspondence, whether of a business or familiar nature (see, for instance, Dossena 2008b, 2010a and 2010b). In particular, as regards the latter, special attention has been paid to the letters encoded by emigrants, in an attempt to provide first-hand material for the study of "language history from below". This is consistent with recent socio-historical lines of research that focus on documents previously disregarded by linguists, on the assumption that they had little, if anything, to offer, because of their supposed divergence from standard texts. In fact, the case is much more complex: encoders often attempted to imitate standard models (for instance, when employing opening and valedictory

formulae), but their texts also provide invaluable instances of vernacular usage (see Schneider – Montgomery 2001).

2. Instances of vernacular syntax and morphology in 19CSC

In this section I intend to focus on two aspects that have often taken centre stage in studies of Scots and Northern English. In particular, I mean to discuss features of syntax and morphology that have typically been assumed to be among the most marked ones in these varieties, i.e. certain uses of modal auxiliaries and the employment of the so-called Northern Subject Rule. Such features will be analyzed in terms of their semantic value and, from the morpho-syntactic point of view, in terms of their relationship with subject type, so as to highlight any constraints that may seem particularly relevant.

2.1 The Northern Subject Rule

This feature, dating back to Older Scots and observed in Northern Middle English (see King 1997 and Macafee 2002), is normally described as the rule according to which the Standard English contrast between the verbal endings *-s* in the 3rd person (singular) and *-ø* in the other persons (singular and plural) only occurs when the adjacent subject is a personal pronoun; when it is not, the ending *-s* may (and often does) occur in place of the ending *-ø* in forms other than 3rd person singular ones. Instances of this feature are frequent in proverbs and sayings, such as "Bannocks is better nor nae bread" or "When the kye comes hame". Murray (1873) discussed its historical roots quite extensively in *The Dialect of the Southern Counties of Scotland*:

> The modern Scotch usage, thay cum, the men cums, is identical with that of the Northern Dialect from the 13th century, which is incorrectly said by many English scholars (Mr. Guest, I think, is the father of the mistake), to have made all the persons of the present tense in -s. But this was only when the pronoun subject was absent; when accompanied by the pronoun, this tense was inflected (with exception of 2nd pers. sing. in -es, thow loves), as in modern literary English. In the Old North-Anglian indeed, the conjugation was:
>
> | | |
> |---|---|
> | Ih cyme | we cym-es |
> | ðu cym-es | ȝee cym-es |
> | he cym-es | hea or þa cym-es […] |

> But before the date of the earliest Northern writings of the 13th century, the form without the -s had been extended to all cases in which the verb was accompanied by its proper pronoun, whether before or after it, leaving the full form in -s to be used with other nominatives only. […] In the verb BE where the plural (aron, area, are, ar, er, yr) did not end in -es, the presence or absence of the pronoun subject did not affect the form of the verb originally; but at a later date, the analogy of the other verbs, in which a form identical with the 3rd pers. sing. was used in the plural in the absence of the pronoun, led to the use of es, is, in like cases for ar, er, though only as an alternative form. In the same way was, wes, intruded upon wer, war, in the past tense. (Murray 1873, in CMSW)

In much more recent times, the possible origins of this phenomenon have been discussed in various studies, most notably by Klemola (2000) and Pietsch (2005), while Ramisch (2010) and Fernández-Cuesta (2011) have highlighted its occurrence in English dialects and with first-person subjects. Indeed, the relative diffusion of this feature was described as late as 1915 in James Wilson's *Lowland Scotch as Spoken in the Lower Strathearn District of Perthshire*:

> As regards number and person, the chief difference between S. and E. [Scots and English] is that in the present tense the ending in z or s, which marks the third person singular, is in S. often used in all persons of the plural unless the verb follows immediately after a single pronoun, and also in the first person singular, especially when the present is used for a narrative past. (Wilson, 1915, in CMSW)

Interestingly, however, no occurrences are observed in CMSW, possibly because these texts, typically composed by educated authors, are based on published editions, and editorial interventions may have silently ironed out instances of "obvious mistakes". In 19CSC, instead, only manuscript sources are included, and several such instances are recorded, a few examples of which are given below. An occasion of existential usage is offered in (1), while (2) and (3) appear to indicate a collective subject; finally, (4) and (5) show the different forms of the verb with pronominal and nominal subjects in one and the same sentence:

(1) I intend to go to Melbourne for there *is* good wages there. (East Strathdownie, 08.04.1856)

(2) I see by your letter that wages *seems* to be on the increase in the Highlands this year. (Liddle, 08.03.1890)

(3) I beleive [sic] times *is* pretty dull all over at present. (Gallup, 28.03.1890)

(4) I hope all the old friends *is* well I had a letter from Canada yesterday they *are* all in good health. (Liddle, 08.03.1890)

(5) how *is* all the old fellows making out. what *is* Sandy and John Murray doing *are* they still in the fishing. I suppose they *are.* (Granite, 31.10.1890)

Instances of actual, unmonitored usage thus seem to indicate the extent to which this form was probably grammaticalized for these speakers, who do not seem to be aware of the morphological discrepancy; indeed, they use it quite consistently, and never correct themselves. At the same time, these instances seem to stress the importance of the subject-type constraint over the non-proximity constraint (cf. Schneider – Montgomery 2001 and Fernández-Cuesta 2011). Similar patterns of contrast between spontaneous usage and supposedly standard grammatical forms are also found in relation to modal auxiliaries, which are the object of the next section.

2.2 Modality

Both epistemic and deontic modality play a very important role in any text. When we focus on correspondence – and, in particular, on emigrants' letters – we see that this role acquires special prominence. In such letters, in addition to information about self and family and descriptions of people and places, emigrants typically included personal opinions and evaluations concerning their current life, or that of their recipients, and expressed requests and wishes, hopes or intentions for the future.

The frequency of modal auxiliaries in this section of 19CSC, given in Table 1 below, appears to reflect typically Scots usage, with few occurrences of *shall*, and relatively rare uses of *might, could* and *should*. However, it should be noted that this part of CMSW comprises ca. 513,000 words, and that nearly 400,000 of these come from just three texts by educated speakers (Elizabeth Grant's *Memoirs of a Highland Lady*, 1898; John Miller's *Diary*, 1889; and Elizabeth Spence's *Letters from the North Highlands*, 1817), which probably

accounts for the much higher frequencies of *shall*, *should* and *might* here than in 19CSC. As for the numerous occurrences of *would* in CMSW, these may be explained in terms of the frequent use of reported speech in memoirs and diaries.

Table 1. Modal auxiliaries in the corpora under investigation (raw numbers and, in italics, normalized figures per 10,000 words)

	Can	*Could*	*May*	*Might*	*Must*	*Shall*	*Will*	*Should*	*Would*
19CSC	49	12	74	8	52	7	230	17	71
19CSC	*4,08*	*1,00*	*6,17*	*0,67*	*4,33*	*0,58*	*19,17*	*1,42*	*5,92*
CMSW (19C personal writing)	279	28	324	193	236	114	361	260	736
CMSW	*5,44*	*0,55*	*6,32*	*3,76*	*4,60*	*2,22*	*7,04*	*5,07*	*14,35*

In 19CSC we find 51 occurrences of *I will* and 4 of *we will*, as opposed to 62 of *you will*, which indicates that the rule recommending *shall* with first-person subjects was not followed in these letters, despite the indications of prescriptive literature that had been so popular since the sixteenth century, and that had made this rule a frequent shibboleth, occasionally with humorous overtones (even David Hume had included notes on modal usage in his 1752 list of Scotticisms: see Dossena 2005: 66). In 19CSC a similar frequency pattern is observed in the case of *would*, to which the same rule prescribing subject combination applied, and which occurs 18 times with *I*, 2 with *we* and 11 with *you*.

In CMSW *I will* is actually more frequent than *you will*, while *we shall* is more frequent than *we will*, though of course the influence of text type conventions cannot be excluded when relatively heterogeneous documents are taken into consideration.[2] Nor should we exclude the importance of the encoders' writing ability in this respect: McCafferty – Amador Moreno (forthcoming) have recently highlighted uses of *will* with first-person subjects as instances of 'change from below' in their investigation of a corpus of Irish correspondence, and, together with other comparable studies of modality in Late Modern varieties, shed new light on the development of regional Englishes both in Britain and overseas. Indeed, even formal letters

2 As the current online interface does not allow searching for phrases, I wish to thank the compilers of CMSW for permission to access the full corpus as text files.

may include vernacular uses that reflect the encoder's own grammatical competence, such as in "sir you gave a good advice but I am afraid you will not can take it yourself" (19CSC).

Concerning the semantic value of modal auxiliaries, in both corpora *may* and *must* are seen to occur with deontic as well as epistemic values – see the examples below: (6, 7, 8, 9) and (10, 11, 12, 13) respectively.

(6) hope he may recover. (19CSC)

(7) I will make few or no alterations, in order that there may be no delay. (CMSW)

(8) you must excuse me at present for my short letter. (19CSC)

(9) you must not think I forget my friends at home. (CMSW)

(10) I may pull up into Colorado. (19CSC)

(11) You may remember that Cairoli in a measure saved the Kings life while riding through Naples a short time ago. (CMSW)

(12) It is sad news indeed about Willie and it must have been a very crushing thing for him, poor fellow. (19CSC)

(13) These visits must have been regarded as formidable undertakings. (CMSW)

Epistemic modality is also employed metacommunicatively when encoders describe the reality in which recipients are expected to find themselves; in 19CSC uses of *suppose* (17 occurrences, all with first-person subjects) and *guess* (8 occurrences, all with first-person subjects too) typically focus on the recipient, as in (14) and (15) below:

(14) I suppose you will be having your Social Gathering now. (19CSC)

(15) I guess you will be having nice weather now. (19CSC)

In such instances the identification with the recipients is further emphasized by the use of the progressive form: while this is more frequent in Scots syntax

than in Southern English (see Miller 1993), it is nonetheless interesting to see it occur in contexts where the virtual simultaneity of action on the part of the recipient and of thought on the part of the encoder creates the imagined reality conveyed in the letter.

As for deonticity, this is expressed in the form of intentions, requests, hopes, advice and promises; while intentions and requests centre on the encoder, hopes, advice and promises generally focus on the recipient, as the vocabulary encoding them communicates what kind of reality the encoder devises for them.

In addition to its articulation in modal auxiliaries, deontic modality is expressed by means of lexical items that need to be tagged accurately in order to avoid ambiguities: the verb *to want*, for instance, is used in the sense of *lack* in a promise, as in (16):

(16) as long as I live and my properity prospers you will not want a little Present every year. (19CSC)

This reflects Older Scots usage: in the *Dictionary of the Scots Language* the following example is provided among others in the entry for *want, v.*: "The thesaurar to sers & deliver all weychts that wantis" (1577, Haddington Burgh Records). But it also features in a direct, bald-on-record request in a much later text:

(17) I want you to bring 5 of the best kind of watches. (19CSC)

The expression of deonticity, on the other hand, does not merely concern requests. In familiar letters, and all the more so in emigrants' letters, hopes for the future or in fact even for the present of both encoders and recipients are always a very important part of the message, as the expression of emotional participation in each other's lives signals proximity and is crucial for the maintenance of social bonds. In the 19CSC section comprising emigrants' letters, the verb *to hope* is seen to occur as many as 98 times: on average, more than twice per letter. The phrases in which it occurs are always encoder-oriented (in *I / we hope, let us hope, hoping*) and recipient-directed (*hoping this finds you well / you will answer soon*). The verb phrase is occasionally reinforced with adverbs like *sincerely* and *earnestly*, both of which boost the truthfulness of the statement, beyond formulaic and predictable usage.

On the other hand, the encoder presents his statements much more modestly when advice is given; in such cases we come across two occurrences

of *my advice* and three of the lexical verb *to advise*. In fact, both noun and verb may co-occur in the same sentence (18), showing that the encoder did not pause to correct himself and avoid the repetition – style was clearly not a primary concern:

(18) he asks my advice in the matter and I have advised him not to enter into aney despewit with you. (19CSC)

Other lexical items indicating both deonticity and awareness of face-saving requirements, such as *suggest, recommend* and *counsel*, never occur in 19CSC, possibly on account of their more formal overtones. In CMSW we have 11 instances of *suggest*, 24 of *recommend*, and 3 of *counsel, v*.

Finally, it is interesting to observe how 19CSC encoders express mutual trust, on the basis of which advice may be given and received. Despite occasional differences with what is observed in the business correspondence section of the same corpus, in familiar letters modesty and mutual esteem are conveyed by means of linguistic strategies that also highlight common ground. In particular, faith appears to play a very significant role, as the expression of trust is typically employed with reference to trust in God or Providence, and not to the need to establish mutual credibility, thanks to the closeness of the relationship between participants (see Dossena, in preparation).

3. Concluding remarks

This study, albeit brief and restricted to very few instances of vernacular usage, has nonetheless shown the importance of studying unedited manuscripts for the investigation of spontaneous usage. While the documents in CMSW are mostly instances of educated usage, in which relatively few (if any) features of Scots blend into otherwise fully English texts, the texts in 19CSC allow us to investigate syntax and morphology in letters and diaries written by people from a variety of social conditions. When we see that CMSW comprises numerous documents by Sir Walter Scott, James Hogg, David Livingston, and other leading figures in the Scottish cultural world of the time, comparing occurrences with those in a corpus comprising documents by far less well-known and (mostly) far less educated people becomes hardly practicable, requiring caveats like those which have been outlined in this study.

In particular, the 19CSC section including emigrants' correspondence features numerous texts encoded by people whose level of education did not enable them to write fully standardized English. What they wrote, instead, reflected their linguistic competence not only as far as spelling is concerned but also – perhaps even more importantly – in relation to syntax and morphology.

We therefore witness totally unselfconscious occurrences of the Northern Subject Rule, and modal choices that reflect long- and well-established Scots patterns. In these letters codes do more than mix: syntactic forms are seen to blend, and a special kind of grammaticalization appears to be at work in the constructions of users who command both Scots and English, though each to a varying extent. Indeed, such documents present the kind of linguistic cline that has been a feature of Scottish usage for many centuries – at least since the days when early codifiers of the English language began to recommend southern models for written discourse while speakers preserved their own usage in everyday exchanges (see Dossena 2005). In the case of modal auxiliaries, then, the rules that prescriptive grammarians had tried to reinforce since the sixteenth century, i.e. the "Wallis" rules that encouraged users to use *shall/should* with first-person subjects and *will/would* with second- and third-person subjects, are seen to apply only in the texts written by more educated speakers. Further down the social ladder such rules do not seem to have been acquired or applied to any great extent.

Lastly, from the methodological point of view, this study has underlined the importance of being wary of clear-cut distinctions and labels. While "personal writing" may be seen to constitute a set of texts comparable to other corpora of correspondence, the kind of encoders whose texts are included in the corpus, the way in which individual texts account for larger or smaller parts of the corpus itself, and the fact that not all these texts are actually based on authentic manuscripts, but rely on published editions, have a considerable impact on what kind of language instances materialize in the corpus itself.

Only in the past ten years have scholars started to write the history of Late Modern English; in the case of Scots and Scottish English, the road behind is even shorter and the road ahead longer, and longer still when "language history from below" is also meant to be considered. But hopefully small steps will continue to be taken in the right direction. Meanwhile, we should carry on gathering authentic samples in order to enrich the basis on which investigations may be conducted, as well as strive to improve on the analyses from which the process has started.

REFERENCES

Sources

Corpus of Modern Scottish Writing (CMSW) Compiled by John Corbett *et al*. (University of Glasgow). http://www.scottishcorpus.ac.uk/cmsw/.

Corpus of Nineteenth-century Scottish Correspondence (19CSC) Compiled by Marina Dossena and Richard Dury (University of Bergamo). In preparation.

Dictionary of the Scots Language (DSL) http://www.dsl.ac.uk/.

Special studies

Dossena, Marina

2001 "The voice of witnesses in nineteenth-century accounts of the highland clearances", *Review of Scottish Culture* 13, 40-50.

2004 "Towards a corpus of nineteenth-century Scottish correspondence", *Linguistica e Filologia* 18, 195-214.

2005 *Scotticisms in Grammar and Vocabulary*. Edinburgh: John Donald (Birlinn).

2006 "19CSC, ICAMET and the diachronic study of specialized discourse in correspondence". In: C. Mair – R. Heuberger (eds.) *Corpora and the History of English*. Heidelberg: Winter, 65-77.

2008a "Many strange and peculiar affairs: Description, narration and evaluation in Scottish emigrants' letters of the nineteenth century", *Scottish Language* 27, 1-18.

2008b "We beg leave to refer to your decision: Pragmatic traits of nineteenth-century business correspondence". In: M. Dossena – I. Tieken-Boon van Ostade (eds.) *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang, 235-255.

2010a "Building trust through (self-)appraisal in nineteenth-century business correspondence". In: P. Pahta et al. (eds.) *Social Roles and Language Practices in Late Modern English*. Amsterdam: Benjamins, 191-209.

2010b "Be pleased to report expressly: The development of public style English in nineteenth-century business and official correspondence". In: R. Hickey (ed.) *Eighteenth-Century English. Ideology and Change*. Cambridge: Cambridge University Press, 293-308.

Forthcoming a "The study of correspondence: Theoretical and methodological issues". In: M. Dossena – G. Del Lungo Camiciotti (eds) *Letter Writing in Late Modern Europe*. Amsterdam: Benjamins.

Forthcoming b "Colliers, farmers, strikers. The discussion of work-related themes in Late Modern Scottish letters and diaries", *Revista de Lenguas para Fines Específicos* 17. [Special Issue (Autumn 2011): *Diachronic English for Specific Purposes*. Guest editors: F. Alonso-Almeida – S. Marrero--Morales.]

Forthcoming c "I write you these few lines: Metacommunication and pragmatics in nineteenth-century Scottish emigrants' letters". In: A. Hübler – U. Busse (eds.) *The Meta-communicative Lexicon of English Now and Then: A Historical Pragmatics Approach*. Amsterdam: Benjamins.

In preparation "Mixing genres and reinforcing community ties in nineteenth-century Scottish correspond ence: Formality, familiarity and religious discourse". In Joanna Kopaczyk (ed.) *English Communicative Communities: A Historical Look.*

Dossena, Marina – Richard Dury

2008 "19CSC, second-generation corpora and the history of English". In: A. Martelli – V. Pulcini (eds.) *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat*. Monza: Polimetrica, 401-412.

Dury, Richard

2006 "A corpus of nineteenth-century business correspondence: Methodology of transcription". In: M. Dossena – S. M. Fitzmaurice (eds.) *Business and Official Correspondence: Historical Investigations*. Bern: Peter Lang, 193-205.

2008 "Handwriting and the linguistic study of letters". In: M. Dossena – I. Tieken-Boon van Ostade (eds.) *Studies in Late Modern English Correspondence: Methodology and Data*. Bern: Peter Lang, 113-135.

Fernández-Cuesta, Julia

2011 "The Northern Subject Rule in first-person-singular contexts in Early Modern English", *Folia Linguistica Historica* 32, 89-114.

Jucker, Andreas H. – Irma Taavitsainen (eds.)

2008 *Speech Acts in the History of English*. Amsterdam: Benjamins.

King, Anne

1997 "The inflectional morphology of Older Scots". In: C. Jones (ed.) *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press, 156-181.

Klemola, Juhani

2000 "The origins of the Northern Subject Rule: A case of early contact?" In: H. Tristram (ed.) *Celtic Englishes II*. Heidelberg: Winter, 329-346.

Kohnen, Thomas

2007 "Text types and the methodology of diachronic speech act analysis". In: S. M. Fitzmaurice – I. Taavitsainen (eds.) *Methods in Historical Pragmatics*. Berlin: de Gruyter, 139-166.

2008 "Tracing directives through text and time: Towards a methodology of a corpus-based diachronic speech-act analysis". In: A. H. Jucker –

I. Taavitsainen (eds.) *Speech Acts in the History of English*. Amsterdam: Benjamins, 295-310.

Lass, Roger
2004 "Ut custodiant litteras: Editions, corpora and witnesshood". In: M. Dossena – R. Lass (eds.) *Methods and Data in English Historical Dialectology*. Bern: Peter Lang, 21-50.

Macafee, Caroline
2002 "A history of Scots to 1700". In: M. Dareau et al. (eds.) *A Dictionary of the Older Scottish Tongue*. Oxford: Oxford University press, xxix-clvii; online at www.dsl.ac.uk/, Section 7.8.2 (accessed November 2011).

Mäkinen, Martti
2006 "Dissemination of textual material and second-generation corpora: Finding intertexts in Early English medical writing". In: M. Dossena – I. Taavitsainen (eds.) *Diachronic Perspectives on Domain-Specific English*. Bern: Peter Lang, 191-205.

McCafferty, Kevin – Carolina Amador-Moreno
Forthcoming "I will be expecting a letter from you before this reaches you. A corpus-based study of *shall/will* variation in Irish English correspondence". In: M. Dossena – G. Del Lungo Camiciotti (eds) *Letter Writing in Late Modern Europe*. Amsterdam: Benjamins.

Miller, Jim
1993 "The grammar of Scottish English". In: J. Milroy – L. Milroy (eds.) *Real English: The Grammar of English Dialects in the British Isles*. London: Longman, 99-138.

Pietsch, Lukas
2005 *Variable Grammars: Verbal Agreement in Northern Dialects of English*. Tübingen: Niemeyer.

Ramisch, Heinrich
2010 "Analysing present-tense verb forms in the *Survey of English Dialects* and the *English Dialect Dictionary*". In: M. Markus et al. (eds.) *Joseph Wright's English Dialect Dictionary and Beyond. Studies in Late Modern English Dialectology*. Bern: Peter Lang, 239-248.

Schneider, Edgar W. – Michael B. Montgomery
2001 "On the trail of early nonstandard grammar: An electronic corpus of Southern U.S. antebellum overseers' letters". *American Speech* 76(4), 388-410.

Tourism as a specialised discourse: The case of normative guidelines in the European Union

Stefania Maci

University of Bergamo

ABSTRACT

Amongst the genres characterizing the language of tourism, some make it seems more immediately recognizable as specialized discourse. These are the genres of normative guidelines related to the tourism industry and issued by the European Union. Given the lack of analysis of these texts from the applied linguistics perspective, this preliminary investigation aims at describing some of the features of such normative guidelines, i.e., the high frequency of complex sentences, modality, technical terminology and ancient crystallized legal forms. Overall, these features seem an expression of statutory language which explicates the necessity to include all possible information in order to avoid ambiguity and vagueness, and to convey an impression of semantic objectivity and legal impartiality.

1. Introduction

Until the 1990s, European governments felt their role was unnecessary in the tourism industry: as long as tourism growth was recorded, the trend in the tourism industry and in communications was characterized by a *laissez faire, laissez passer* policy. This tendency was shared to such an extent that many Ministries of Tourism ceased to exist or were incorporated into other ministries, and although national tourist boards were established, they had a mere promotional role. At the beginning of the 21st century, new economies emerged and supply exceeded demand: tourists had endless choices, which substantially increased competition in the tourism market. New political issues, in particular public policies, were needed, since they were regarded as

being of vital importance because of their influence in the tourism industry and the regulation of related activities (Klancnik 2006: 58).

The important role of tourism in the European economy has been recognized by the European Union (EU) ever since the early 1980s. For instance, the EU's major contribution to the tourism industry, *Communication on a renewed tourism EU policy: towards a stronger partnership for European Tourism* (COM (2006) 134 final) of 17.03.2006.), issued in 2006, is an attempt to deal with the challenges of the 21st century by taking advantage of all resources and synergies present in the EU states, aiming at promoting better regulations at a European level, given the large number of different policies affecting tourism, in order to (a) minimize the negative impacts of tourism on society/the environment and (b) maximize tourism's positive and creative contribution to local economies. Since then, attention has become more and more focussed on tourism from a normative viewpoint. The consequence, at a linguistic level and from a professional perspective, is that legal discourse has affected the language of tourism to such an extent that expert members of the tourist discourse community tend to adopt lexico-grammatical, rhetorical and discoursal generic conventions deriving from legal discourse, which results in the "colonization of one genre by the other by invading its integrity" (Bhatia 2004: 87). In particular, the use of legal features in tourism discourse and genres seems to suggest that tourism is inclined towards the acquisition of aspects whose function is strictly related to normative and authoritative discourse in order to convey a more ruled image of the tourist community "in adverse and challenging economic circumstances" (Bhatia 2007: 395).

Although the unprecedented increase in global tourism has created zones of mixed jurisdiction (Bhatia 2003: 353), and despite the fact that the EU covers both common law and civil law systems, the EU legislation regulates member states mainly – but not necessarily – as regards the financial sector, to which all member states have to conform[1] (http://europa.eu/abc/treaties/index_en.htm), whereas in the social, defensive, and educational fields, each member follows its own regulations. In the case of tourism, the EU does not set down any real legislation, but rather provides guidelines which each member state applies according to what legal system the country has adopted, i.e. civil law or common law.[2] This implies that the type of legal system in force in a country is not relevant in tourist texts.

---

1 The EU policy requires that: "laws (regulations, directives and decisions) take precedence over national law and are binding on national authorities" (http://ec.europa.eu/community law/introduction/treaty_en.htm).

2 This is particularly true for such countries as the United Kingdom, Poland and the Czech Republic which, according to the Lisbon Treaty in force as of December 1st,

From the applied linguistics perspective, the language of tourism has been the focus of many investigations, mainly analyses of the language of tourism promotion (Calvi 2000, 2001, 2006; Cappelli 2006; Francesconi, 2005, 2007; Gotti 2006; Catenaccio 2009, Maci 2010), and tourism mediation (Nigro 2006; Cappelli 2007), also investigated from a corpus-based perspective (Tognini Bonelli – Manca 2002). On the other hand, legal language has been extensively analysed, with particular regard to the analysis of metatextual markers (Bhatia 1987), of questioning (Pascual 2002, 2006; Sala 2010) and from both a forensic perspective (Cotteril 2003; Gibbson 2003) as well an ethnographic viewpoint (Walter 1988). Attention has also been paid to the generic tension of legal texts across cultures (Bhatia 2005; Gotti 2005), to the construction of professional legal identities (Sala 2008) and to the use of legal language (Hiltunen 1990; McMeel 2005; Williams 2005) in more general terms.

To the best of my knowledge, there is no study concerning the analysis of EU normative language applied to the tourism industry. Keeping in mind that EU normative texts are regarded as guidelines to be locally 'translated' by each member state, this paper tries to describe the type of language characterizing such texts related to tourism in order to see if and to what extent the languages of legal and tourism discourse are interrelated.

The preliminary results of this investigation will show that European normative guidelines follow the generic conventions and constraints of legal discourse while presenting an extensive set of items characteristic of the tourism industry.

2. Legal language in tourism: The corpus

In order to understand in what ways normative and tourism discourses are interrelated, all legal documents concerning tourism were collected by the Euro-lex search-engine (http://eur-lex.europa.eu/RECH_menu.do?ihmlang=en), the EU website granting access to European Union Law. A search was carried out using the keywords 'tourism' and 'tourist'. All legal documents and relative annexes issued between 1984 and 2009 were found and downloaded. All texts were digitalized in a text format, forming a corpus

2009, can opt to follow EU regulations or not (http://europa.eu/abc/treaties/index_en.htm). My thanks to Prof. Cosimo Notarstefano for his explanation regarding the relationship between European legislation and the member states' legal systems.

of 49,459 words; a quantitative computation with Wordsmith Tools (Scott 2007) was then followed by a qualitative analysis. Figures, when given, have been normalised per thousand words and are expressed as Type/Token Ratio (TTR). Full reference to legal texts, from which the excerpts presented in the following paragraphs are taken and here indicated in increasing order, may be seen in the Appendix.

3. Results: Legal discourse in the tourism industry

Literature on legal language (Bowers 1989; Solan 1993; Gibbons 1994; Tiersma 1999; Gotti 2003: 42-46) suggests that the type of texts forming legal genres exploits a type of discourse very different from everyday speech, especially when legal texts express regulations creating, modifying or terminating the rights of and obligations towards individuals and institutions. Such textual differences are evident both at a macro- and at a micro-level, respectively featured by extremely long sentences and cross-references on the one hand, and the use of archaic terms and Latin expressions, binomials and doublets, specialized lexis, nominalization and modality on the other.

3.1 Macro-level analysis: Sentences and cross-references

The only studies dealing with sentence length in the English legal system (Hiltunen 2001; 1985) indicate that sentences have an average of about 45.04 words per sentence. My normative guideline corpus related to tourism comprises 1,489 sentences and has an average of 33.22 words per sentence, which is lower than indicated by Hiltunen (2001; 1985). My computation seems more in keeping with the traditional legal sentence, which, in the 1990s, was measured at 37.06 words per sentence, offering therefore greater readability (Belotti 2002: 115).

In my corpus, any sentence ending with a full stop usually corresponds to a legal topic expressed in a paragraph which contains more than one sentence separated by a series of semi-colons. This atypical orthographic device seems to substitute the conventional full stop used in standard punctuation. If this is so, and we add the number of sentences separated by a semicolon to the general figures given above, the resulting number of sentences will be 1,788; each phrase seems then formed by an average of 27.51 words (see Table 1), which is well below the figures given both by Hiltunen (2001, 1985) and Belotti (2002):

Table 1. Average sentence length

Sentences separated by full stop	1,489
Number of words	49,459
average proposition length	33.22
propositions separated by semi-colon	299
average proposition length	27.51

However, as in my corpus semi-colons do not visually break the sentence, the impression is that of sentences coinciding with very long paragraphs, which obviously has a great impact on readability: the longer the sentence, the more difficult it is to understand. The first analysis involving sentence length has therefore been a classification of sentence types. Simple sentences (containing one predicate) are easily understood because of the simplicity of their syntactic construction. Complex sentences (consisting of one main clause and one or more subordinates), on the contrary, are, indeed, difficult to understand not only because of their length but also because of the presence of embedded clauses, which can be either left- or right-dislocated. Indeed, the presence of subordinates, present participles or gerunds that interrupt or even invert the English structure S + V + O + Adjunct, creates a certain lack of readability:

(1) When constructing electricity transmission lines and power stations linked to them, as well as oil and gas pipelines, including pumping stations and booster stations and plants which are very significant from an environmental point of view, the Contracting Parties shall implement all the necessary measures to avoid disturbance to the local people and the environment, including, if possible, the use of pre-existing facilities and grids.
(EU007)

In the excerpt above, the main clause is preceded by five secondary clauses and is followed by another three secondary ones: overall, we have four relatives, one rendered with the *wh-* relative pronoun, two by means of the present participle form in *-ing*, and one by the past participle ending with *-ed*, one *if*-clause, one purpose clause and one adverbial clause, which impedes readability.

Clause distribution of the collected corpus can be seen in Table 2.

Table 2. Breakdown of simple and complex sentences in the EU tourism guidelines

Simple clauses			1,870
Complex clauses	Secondary clauses introduced by *that*		310
	If-clauses		73
	Relative clauses	*Who*; *which*; *that*	304
		past participle	1,026
		-ing forms	1,491
	Concessives clauses (introduced by: *nevertheless*; *yet*; *however*; *despite*, (*al*)*though*, *even though/if*)		85
	Purpose clauses		1,242

In my corpus, there is a great occurrence of complex sentences in EU tourism normative guidelines: they are concentrated in the definition of the exceptions to the provisions or in the explanation of the provisions, normally annexed to the normative text:

(2) 1.1.1. Collective tourist accommodation establishments Definition:
An accommodation establishment that provides overnight lodging for the traveller in a room or some other unit […]
(EU5a)

This indicates that the main aim of the author is to make less ambiguous any possibly vague meaning in order to provide a basis for an objective and impartial provision of the rules constructed *contra proferentem* (McMeel 2005), i.e., against the interests of the party responsible for drafting the guidelines. In addition, removing any ambiguity implies that all EU guidelines have to be compiled using a detailed and unequivocal description of all legally relevant elements to show an accurate account of the legislator's reasoning. This may be responsible for lengthy sentences.

As illustrated by Hiltunen (2001: 61), *that*- and relative clauses are the most frequently exploited clauses in legal discourse. All relative clauses in my corpus are defining and occur whenever there is a need to unambiguously limit and define facts, people and norms. This clearly fosters clarity of expression, even though it may create a lack of readability:

(3) The expenditure declarations presented to the Commission are not always reliable. In the case of Italy, […]. This practice, which the

Commission accepts, is not in accordance with the requirements set out in its initial decision of 14 December 1990 approving the Italian tourism OP, whereby the expenditure submitted must correspond to the expenditure incurred.
(EU20)

As said above, these definition or explanation sections are in the *annex* to the guidelines which occurs *after* the EU guideline. It is, to be precise, in the statement of the guideline, i.e. before the expansion of the provisions in the annex. The provisions, on the other hand, are linguistically constructed with simple sentences, characterized by a more linear structure and the use of the present simple indicative, the latter implicitly indicating the performative function of the written text (Garzone 2003: 206) because of its directness and unconditional nature. Since simple sentences are usually found whenever the decision or one point in argumentation is to be made clear, they are likely to help render the EU directives less complicated for laypersons.

Lack of readability in legal discourse is also caused by the fragmentation of the syntactical structure of texts because of continuous cross-referencing to laws, regulations, and provisions. Although this feature also characterizes my corpus, the reference to law regulations is normally made by use of endnotes, so such provisions do not interfere with textual cohesion:

(4) Having regarded
  (1) The Convention on the protection of the Alps (hereinafter "the Alpine Convention") was concluded on behalf of the European Community by Council Decision 96/191/EC [2].
  (2) The Council decided on the signature, on behalf of the European Community […], by Council Decision 2005/923/EC [3].
  (EU08)

Interference with textual cohesion is, on the other hand, created by a feature commonly found in my corpus: if sentences correspond to a relevant legal issue, they are generally introduced by numbered and lettered paragraphs:

(5) 3. They shall adopt measures and make provisions, particularly in the following areas:
  (a) improving insulation in buildings and the efficiency of heating systems;

(b) optimising the performance of heating, ventilation and air conditioning systems;
(c) […];
(d) […].
Article 6
Renewable energy resources
1. The Contracting Parties shall energy take […].
2. They shall also encourage the use of […].
(EU19)

Although it interrupts cohesion, this practice seems to facilitate readability because the numbered paragraphs are to be seen as headings in a text concerning a legal topic. Nevertheless, a text so schematically organized, and apparently easier to understand from a semantic viewpoint, is actually denser and more cognitively demanding because of the frequent use of nominalised forms in pre-modifying position (my emphasis):

(6) 1. Consular offices of Member States in the PRC should issue an *accreditation* certificate for each designated *travel* agency.
(EU04)

3.2 Micro-level analysis

3.2.1 Lexis

The analysis of EU guidelines related to tourism points to the occurrence of a legal register characterised by technical terminology, archaic expressions, Latin terms and binomials; the latter, in particular, are outdated and crystallized expressions which mark the permanent nature of statutory provisions (Giannoni 2003). All these elements, together with the massive presence of specialized lexical items and of particular syntactical structures, permeate the language of EU guidelines related to tourism with *legalese*.

Yet the keyword list created with Wordsmith Tools shows the massive presence of terms that apparently do not belong to *legalese* (Fig. 1).

KEYWORDLISTeuropean law.kws							
N	*Key word*	*Freq.*	*%*	*Freq.*	*RC. %*	*Keyness*	*P*
1	TOURISM	378	0.69	1,461		3,813.01	000000
2	SHALL	303	0.56	19,817	0.02	1,430.03	000000
3	ARTICLE	218	0.40	6,607		1,351.76	000000
4	COMMISSION	222	0.41	9,844		1,213.29	000000
5	MEMBER	256	0.47	17,230	0.02	1,193.97	000000
6	STATES	192	0.35	17,873	0.02	776.92	000000
7	TOURIST	107	0.20	1,986		764.65	000000
8	ACCOMMODATION	128	0.23	4,373		763.83	000000
9	COMMUNITY	202	0.37	22,542	0.02	748.60	000000
10	DIRECTIVE	100	0.18	1,718		729.51	000000
11	ESTABLISHMENTS	85	0.16	833		711.27	000000
12	WHEREAS	132	0.24	6,169		707.83	000000
13	NR	57	0.10	97		653.27	000000
14	MEASURES	128	0.23	6,878		651.97	000000
15	CONTRACTING	78	0.14	787		648.37	000000
16	EUROPEAN	162	0.30	20,245	0.02	566.11	000000
17	PROTOCOL	72	0.13	1,032		550.27	000000
18	ALPINE	57	0.10	478		493.70	000000
19	TRAVEL	101	0 19	7,145		461.18	000000
20	THE	4,557	8.36	55,105	6.09	443.01	000000
21	NON	61	0.11	1,241		425.08	000000
22	INFORMATION	173	0.32	38,362	0.04	425.06	000000
23	HOTELS	69	0.13	2,329		413.29	000000
24	ANNEX	42	0.08	198		408.46	000000
25	CONTRACT	106	0.19	11,875	0.01	391.90	000000
26	CONSULAR	37	0.07	116		386.58	000000
27	CONCERNING	70	0.13	3,354		371.93	000000
28	OF	2,497	4.58	49,564	3.07	366.07	000000
29	ACCORDANCE	61	0.11	2,092		363.49	000000
30	DATA	117	0.21	18,084	0.02	363.00	000000

Figure 1. Keyword list of the EU legislation on tourism

Tourism is evidently the strongest keynote term, as it is the topic dealt with in the normative guidelines. The most frequent collocates of *tourism* are shown in Fig. 2 below:

TOURISM.cnc								
N	*Word*	*With*	*Relation*	*Total*	*Total Left*	*Total Right*	*L5*	*L4*
1	TOURISM	Tourism	0.000	420	21	21	6	5
2	THE	Tourism	0.000	325	204	121	28	50
3	OF	Tourism	0.000	240	175	65	19	19
4	AND	Tourism	0.000	125	65	60	11	20
5	IN	Tourism	0.000	115	72	43	4	29
6	TO	Tourism	0.000	87	41	46	11	8
7	ON	Tourism	0.000	68	63	5	13	2
8	FOR	Tourism	0.000	53	41	12	8	1
9	COMMUNITY	Tourism	0.000	52	42	10	7	13
10	A	Tourism	0.000	43	20	23	3	13
11	DOMESTIC	Tourism	0.000	33	19	14	2	2
12	WHEREAS	Tourism	0.000	27	9	18	2	0
13	EUROPEAN	Tourism	0.000	27	24	3	6	1
14	FIELD	Tourism	0.000	26	25	1	0	0
15	IS	Tourism	0.000	26	6	20	4	2
16	MEASURES	Tourism	0.000	22	19	3	5	9
17	INFORMATION	Tourism	0.000	22	19	3	10	0
18	OUTBOUND	Tourism	0.000	21	11	10	0	1
19	BY	Tourism	0.000	21	7	14	2	0
20	DEMAND	Tourism	0.000	19	3	16	0	3
21	FAVOUR	Tourism	0.000	19	19	0	0	0
22	NATIONAL	Tourism	0.000	19	17	2.	1	0
23	STATISTICS	Tourism	0.000	18	4	14	0	0
24	INBOUND	Tourism	0.000	17	13	4	0	2
25	1	Tourism	0.000	17	1	16	1	0
26	BE	Tourism	0.000	17	5	12	2	2
27	AS	Tourism	0.000	17	6	11	1	0
23	WHICH	Tourism	0.000	15	5	10	1	1

Figure 2. Collocates of *tourism*

An analysis of the collocates as well as the clusters of *tourism* reveals that the terms are more in line with tourism planning and management than with legal discourse. Yet all the other keywords clearly point to the legal nature of these guidelines. For instance, *shall* is a top keyword, which is not surprising: the modal *shall* normally carries out a deontic function and indeed, it is used for conveying orders and instructions, and for signalling juridical obligation. Among the first thirty items having a strong keyness, we can find *article*, *commission*, *member states*, *directive* and *protocol* (raw frequency and percentage of the items can be seen in the second and third column of Figure 1). These are specialized technical terms which belong to a legal register. Therefore they do not distinguish EU normative guidelines, in which they appear, from any other type of legal texts. On the contrary, the most distinguishing feature of these normative texts seems to be the occurrence of terms belonging to the tourism industry.

3.2.1.1 Doublets, binomials, archaisms and technical terminology

Doublets and binomials are common in *legalese* and are used to highlight alternative options (Bhatia et al. 2003). The exploitation of binomials and doublets can be traced back to the Anglo-Saxon period characterized by an extensive use of alliteration in legal language. Such doubling continued in medieval English legal practice, which involved the pairing of a native English word placed before an equivalent French word. The main reason for the continuation of this linguistic tradition in English legal documents is that such word-strings are used to convey all-inclusiveness, that is, to cover all possible situations and eventualities, which accounts for their redundancy and wordiness (Cao 2007). This is true for my corpus, where doublets and binomials seem to encode inclusiveness:

(7) *Purpose* and *scope* of application
(EU03a)

(8) short distance local *transport* and *commuting*
(EU05a)

(9) the *budgetary* and *financial* implementation
(EU20)

(10) […] *alerting* and *alarm* system […] *curtains* and *drapes* […] *components* and *materials* […]
(EU12)

Binomials allow for alternatives that, while extending the provision's legal coverage, semantically overlap the meaning of the legal text to allow for contextual interpretation. Member states are thus left free to define the range and scope of that concept with the present provision (Giannoni 2005).

At the same time, there is an extensive presence of technical terminology, such as *commission* (222 hits, TTR 4.44), *tourist* (107 occurrences, TTR 2.14), *protocol* (72 hits, TTR 1.4), *provision* (37 hits, TTR 0.74), *memorandum* (43 hits, TTR 0.86), *annex* (42 hits, TTR 0.84), *retailer* (29 occurrences, TTR 0.58), *purchaser* (28 hits, TTR 0.56), *audit* (21 occurrences, TTR 0.42), *amended* (18 occurrences, TTR 0.36), *beneficiary* (9 hits, TTR 0.18), *traveller* (4 hits, TTR 0.08). Technical terms apparently make the text more semantically complicated but they are necessary so as to determine with great precision the subjects to whom the law is applied.

In addition, the collected corpus features archaic, formal, and at times unusual or difficult vocabulary, such as *whereas* (132 hits, TTR 2.64); *thereof* (26 hits, TTR 0.25), *pursuant* (19 hits, TTR 0.38), *inter alia* (14 hits, TTR 0.28), *hereby* (8 hits, TTR 0.16), *thereby* (4 hits, TTR 0.08), and *force majeure* (2 hits, TTR 0.04). This archaic yet technical register is necessary because it indicates the conditions and situations described in the provision. Such lexical items are generally present in limiting clauses, which are common in English legal documents as they endow a document with a legal tone (Bhatia et al. 2003), but increase sentence length in the process:

(11) THE COMMISSION OF THE EUROPEAN COMMUNITIES, […]
Having regard to Council Directive […] and in particular Articles 3, 7 and 10 thereof,
Whereas, to facilitate […];
Whereas, to facilitate data collection, […];
Whereas, during the transition period, […];
Whereas the measures provided for […]
HAS ADOPTED THIS DECISION.
(EU05)

3.2.3 Modality

Juridical obligation and permission are mainly expressed by modality. As aptly pointed out by certain scholars (Palmer 2001; Gotti et al. 2002, Facchinetti et al. 2004; Ziegeler 2006) modality points to the illocutionary attitude of the speaker as well as the speaker's indications of the probability or necessity of a statement. In particular, modality can be classified as:

– epistemic, when expressing the speaker's degree of commitment to the truth proposition;
– deontic, when expressing the speaker's responsibility/authority in giving permission, imposing commands, offering suggestions;
– dynamic, when expressing the speaker's/listener's ability or disposition.

As we have seen from the keyword list, *shall* appears to be the item carrying the strongest keyword importance in EU normative texts concerning tourism; its raw frequency is 303 (TTR 6.06), and, as Figure 2 below indicates, it is used in the passive form whenever the reference is to the adoption of the provision, and in the active form when the subject of the clause is either the contracting party or the member state to whom the ruling is directed:

SHALL.cnc			
N	*Concordance*		
1	Article 14 Entry into force This Directive	shall	enter into force on the 20th day
2	of the Directive Member States	shall	bring into force the laws
3	and Article 7 of Decision 1999/468/EC	shall	apply having regard to the
4	in Article 4(3) of Decision 1999/468/EC	shall	be set at three months.
5	4 and 7 of Decision 1999/468/EC (2)	shall	apply, having regard to the
6	measures referred to in Articles 7 and 9	shall	be adopted in accordance with the
7	inter alia by supplementing it,	shall	be adopted in accordance with the
3	period (Article 10) the Commission	shall	be assisted in accordance with the
9	during a transition period which	shall	end three years after entry into
10	prejudice to Article 13, Member States	shall	take all the measures necessary to
11	of the data by the Commission	shall	be determined pursuant to the
12	the methods used 2. The Commission	shall	present to the European
13	statistical information. Member States	shall	also provide the Commission with
14	Article 8 Reports 1. Member States	shall	provide the Commission at its
15	revised monthly and quarterly results	shall	be transmitted within a maximum
16	provisional monthly and quarterly data	shall	take place within three months of

17	period, and the revised annual results	shall	be transmitted within a maximum
18	transmission of provisional annual data	shall	take place within six months of the
19	Transmission of data 1. Member States	shall	transmit the data processed in
20	to in Article 12(2). The regional level	shall	be in accordance with the
21	of this Directive by supplementing it,	shall	be adopted in accordance with the

Figure 3. Concordance list of *shall*

Since *shall* is not the only modal expressing deontic obligation, I decided to carry out an analysis focussed on modal verbs in order to investigate expressions stating juridical permission and obligation, and whose results are illustrated in Tables 3 and 4.

Table 3. Modality in the EU guidelines on tourism: *will*, *would*, *shall* and *should*

		Will		*Would*		*Shall*		*Should*	
		Hits	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*
Deontic	*Necessity*	41	0.82	–	–	–	–	112	2,24
	Order	–	–	–	–	303	6.06	–	–
Dynamic	*Necessity*	–	–	24	0.48	–	–	–	–

Table 4. Modality in the EU guidelines on tourism: *may*, *might*, *can*, *could*, *must*, and *need*

		May		*Might*		*Can*		*Could*		*Must*		*Need*	
		Hits	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*	*Hits*	*TTR*
Epistemic	*Probability*	–	–	2	0.04	–	–	–	–	–	–	–	–
	Possibility	41	0.82	3	0.06	11	0.22	17	0.34	–	–	–	–
	Necessity	–	–	–	–	–	–	–	–	105	2.10	7	0.14
Deontic	*Permission*	57	1.15	–	–	30	0.60	–	–	–	–	–	–
	Order	–	–	–	–	–	–	–	–	7	0.14	–	–
Dynamic	*Possibility*	–	–	–	–	–	–	–	–	–	–	–	–
	Ability	–	–	–	–	–	–	–	–	–	–	–	–

The allocation of *shall* (303 hits, TTR 6.06), as we have seen above, indicates the presence of a directive. *Shall*, therefore, has a deontic obligation function:

(12) In the event of the construction of new, large power plants […] the Contracting Parties, in accordance with current law, shall proceed to evaluate the impact on […] in accordance with Article 12. The Parties shall recognise the right to consultation at international level on projects with cross-border effects.
(EU19)

Should (112 hits, TTR 2.24) is used to convey the sense of deontic necessity to suggest that it is necessary to follow some advice, as indicated by example (13):

(13) Trans European Energy Networks (TEN-E) should be given priority and coordination and implementation measures foreseen in the TEN-E guidelines in Decision No 1229/2003/EC of the European Parliament.
(EU08)

The distribution of *will* (41 hits, TTR 0.82) suggests that the modal occurs when it denotes a deontic necessity in conveying instructions:

(14) The Commission will ensure that this task will be completed by June 1997 at the latest; a full report will be provided to the Court at that stage.
(EU20)

Would, on the contrary, is mainly used with a dynamic function (24 hits, TTR 0.48) related to the degree of possibility, probability or impossibility of an action and in reported speech, which seems to have a predictive role:

(15) The stay in the place visited should not last more than 12 consecutive months, beyond which the visitor would become a resident of that place.
(EU05a)

As shown in (16), permission is normally conveyed with the modal *may* (57 hits, TTR 1.15), which also, though less frequently, indicates possibility (41 hits, TTR 0.82). In both cases it is much more frequent than *can.*

(16) […] when the ship is in port, the passengers may or may not be formally free to enter the country.
(EU05a)

The modal *might* seems to occur with a low frequency and only to indicate either epistemic probability or possibility, as we can see from examples (17) and (18) below:

(17) No age limit is applied: children are counted as well as adults, even in the case when the overnight stays of children might be free of charge.
(EU05a)

(18) If the passengers are free to enter the country, the nights might be in principle recorded to that country, […]
(EU05a)

Can is alternatively employed in conveying epistemic possibility (11 hits, TTR 0.22) and deontic permission (30 hits, TTR 0.60):

(19) The destination can be understood in different ways.
(EU05a)

(20) Marinas Definition: Consist of boating harbours where boat owners can hire a berth in the water or a place on the land for the season or year.
(EU05a)

Could is less frequently used than *can* and is employed to indicate the possibility that an action could take place, given the premises (17 hits, TTR 0.34):

(21) […] unforeseeable circumstances beyond the control of the party by whom it is pleaded, the consequences of which could not have been avoided even if all due care had been exercised.
(EU9)

Must is used only in an epistemic sense, indicating the necessary action to be taken in order to exercise the right (105 hits, TTR 2.10):

(22) […] require that the contract be drawn up in its language or its languages which must be an official language or official languages of the Community.
(EU18)

The deontic modality expressed by *must* occurs only in seven cases and in negative expressions only (TTR 0.14) and indicates what is forbidden:

(23) Any descriptive matter concerning a package and supplied by the organizer or the retailer to the consumer, the price of the package and any other conditions applying to the contract must not contain any misleading information.
(EU9)

Need (7 hits, TTR 0.14) is similar to *must*, when indicating the epistemologic necessity to undertake an action in order to implement the right:

(24) Since October 1994 it has however been sent the documents it needs to follow up the case.
(EU20)

There is, however, a prevalence of the use of *need* in its nominalised form (27 hits, TTR 0.54), as excerpt (25) below shows:

(25) The reorganization within the Tourism unit of DG XXIII will take account of the need to increase its involvement in indirect actions in favour of tourism.
(EU20)

Clearly, all these modals are distributed differently within the guidelines: *shall*, *can, may* and *must* are mainly found in the orders section, whereas all the other forms are mainly – but not necessarily – found in the *annex*, i.e., the technical explanations to the guideline.

4. Conclusions

The documents here analyzed offer an example of interdiscursivity in the sense that the EU guidelines devoted to tourism are lexically colonized by legal discourse language.

Amongst the characteristics of the language of legal documents featured in these guidelines, is the high frequency of lengthy and complex sentences. Indeed, statutory language is characterized by this feature, since legislators tend to such an extent to include all possible information

in a single sentence in order to avoid ambiguity and vagueness, that they increase considerably the level of information density. Furthermore, general principles are expressed in such a way as to allow flexibility which, on the one hand, gives the impression of a lack of vagueness (Bhatia 2003: 338), but on the other has to be detailed and specific because of their performative function; hence the use of modals, binomials, technical terminology and archaic crystallized legal forms such as the use of doublets and double negative patterns which, while increasing semantic density, convey an idea of semantic objectivity and an apparent idea of legal impartiality, this in turn suggesting a stronger sense of authoritativeness.

The EU presents its normative texts both on-line and in a traditional paper format. At the same time, it is aware of the urging societal and economic developments influencing tourism. Such texts are primarily legal and follow the generic conventions and constraints of legal discourse while presenting an extensive set of lexical items characteristic of the tourism industry. Legal texts are rich in cross-references to previous norms and provisions which, for a layman, are difficult to interpret because they are seen as not being organized in a cohesively and hierarchically sequential order. Meaning-making in such texts as those analysed in this study is thus a complex process because the domains dealt with are primarily practice-oriented (i.e., aimed at normative administration in the tourism industry). For the purpose of normative effectiveness, EU texts seem to be prone to adopting all those rhetorical linguistic devices typical of legal discourse which enable the specialist reader to better comprehend and assimilate the content, and possibly to turn it into practice.

REFERENCES

Belotti, Ulisse

2002 "The language of Italian arbitration rules in English: Some measurable aspects", *Linguistica e Filologia* 15, 113-141.

Bhatia, Vijay K.

1987 "Textual-mapping in British legislative writing", *World Englishes* 6 (1), 1-10.

2003 "Specificity and generality in legislative expression". In: V.K. Bhatia et al. (eds.) *Legal Discourse in Multilingual and Multicultural Contexts. Arbitration Texts in Europe*. Bern: Peter Lang, 337-356.

2004 *Worlds of Written Discourse: A Genre-based View*. London: Continuum.

2005 "Specificity and generality in legislative expression: Two sides of the same coin". In: V.K. Bhatia et al. (eds.) *Vagueness in Normative Texts*. Bern: Peter Lang, 337-356.

2007 "Interdiscursivity in critical genre analysis", *Proceedings of the 4th Sympósio Internacional de Estudos de Gêneros Textuais*.

http://www3.unisul.br/paginas/ensino/pos/linguagem/cd/English/36i.pdf. Bhatia, Vijay K. et al. (eds.)

2003 *Legal Discourse in Multilingual and Multicultural Contexts*. *Arbitration Texts in Europe*. Bern: Peter Lang.

Bowers, Frederick

1989 *The Linguistic Aspects of Legislative Expression*. Vancouver: University of British Columbia Press.

Calvi, M.V.

2000 *Il linguaggio spagnolo del turismo*. Viareggio: Baroni Editore.

2001 "El léxico del turismo". In: Á.F.M. Lago (ed.) *EIELE para contextos* profesionales. Cultura e intercultura en la enseñanza del español a *extranjeros*. http://www.economiaturismoassisi.it/public/materiale/docenti/53/file/El%20l%C3%A9xico%20del%20turismo.doc.

2006 "El uso de términos culturales en el lenguaje del turismo: los hoteles y su descripcion". In: M.V. Calvi et al. (eds.) *Nuevas Tendencias en el Discurso de Especialidad*. Bern: Peter Lang, 271-292.

Cao, Deborah

2007 *Translating Law*. Clevedon: Multilingual Matters Ltd. Cappelli, Gloria

2006 *Sun, Sea, Sex and the Unspoilt Countryside. How the English Language Makes Tourists out of Readers*. Pari: Pari Publishing.

2007 "The translation of tourism-related websites and localization: Problems and perspectives". In: A. Baicchi (ed.) *Voices on Translation, RILA Rassegna Italiana di Linguistica Applicata*. Roma: Bulzoni Editore, 97-115.

Catenaccio, Paola

2009 "Promoting medical tourism in India: Building identity and reputation in health tourism websites". In G. Garzone et al. (eds.) *Identities Across Media and Modes*. Bern: Peter Lang, 59-108.

Charrow Robert P. – Veda R. Charrow

1979 "Making legal language understandable: A psycholinguistic study of jury instructions", *Columbia Law Review* 79 (7), 1306-1374.

Cotterill, Janet

2003 *Language and Power in Court: A Linguistic Analysis of the O.J. Simpson Trial.* Houndmills: Palgrave Macmillan. Facchinetti, Roberta et al. (eds.)

2004 *English Modality in Perspective*. Frankfurt am Main: Peter Lang.

Francesconi, Sabrina
2005 "The language of souvenirs: The use of humour in London t-shirts", *Textus* 18, 381-396.
2007 *English for Tourism Promotion*. Milano: Hoepli.
Gibbons, John
1994 *Language and the Law*. London: Longman.
2003 *Forensic Linguistics.* Oxford: Blackwell.
Garzone, Giuliana
2003 "Arbitration rules across legal cultures: An intercultural approach". In: V.K. Bhatia et al. (eds.) *Legal Discourse in Multilingual and Multicultural Contexts. Arbitration Texts in Europe*. Bern: Peter Lang, 177-220.
Giannoni, D.S.
2003 "The UNITRACIL Model and Italian statute law: A linguistic and topical description". In: V.K. Bhatia et al. (eds.) *Legal Discourse in Multilingual and Multicultural Contexts.* Bern: Peter Lang, 221-246.
2005 "'Any dispute shall be settled by arbitration'. A study of vagueness in international model arbitration clauses". In: V.K. Bhatia et al. (eds.) *Vagueness in Normative Texts*. Bern: Peter Lang, 437-468.
Gotti, Maurizio
2003 *Specialized Discourse*. Bern: Peter Lang.
2005 "Vagueness in the Model Law". In: V.K. Bhatia et al. (eds.) *Vagueness in Normative Texts.* Bern: Peter Lang, 227-253.
2006 "The language of tourism as specialized discourse". In: O. Palusci et al. (eds.) *Translating Tourism Linguistic / Cultural Representations*. Trento: Editrice Università degli Studi di Trento, 15-34.
Gotti Maurizio et al. (eds.)
2002 *Conflict and Negotiation in Specialized Texts*. Bern: Peter Lang. Halliday, M.A.K. – Christian Matthiassen.
2004 *An Introduction to Functional Grammar*. London: Arnold. Hiltunen, Risto.
1985 "Aspects of structures in some English legal texts". In: J. Tommolaet et al. (eds.) *Working Papers in English Studies*. Turku: University of Turku.
1990 *Chapters on Legal English*. Helsinki: Suomalainen Tiedeakatemia.
2001 "Some syntactic properties of English law language". In: R. Hiltunen et al. (eds.) *English in Zigs and Zags: A Festschrift for Marita Gustafsson*. *Special Issue of Anglicana Turkuensia* 23, 53-66.
Klancnik, R.V.
2006 "Media and communications". In: D. Buhalis et al. (eds.) *Tourism Management Dynamics. Trends, Management and Tools.* Amsterdam: Elsevier, 58-67.
Maci, Stefania M.
2010 *The Language of Tourism*. Bergamo: CELSB.

McMeel, Gerard
2005 "Language and the law revisited: An intellectual history of contractual interpretation", *Common Law World Review* 34 (3), 256-286.
Nigro, M.G.
2006 *Il linguaggio specialistico del turismo. Aspetti storici, teorici e traduttivi*. Roma: Aracne.
Palmer, R.F.
2001 *Mood and Modality*. Cambridge: Cambridge University Press.
Pascual, Esther
2002 *Imaginary Trialogues: Conceptual Blending and Fictive Interaction in Criminal Courts.* (LOT Dissertation Series 68.) Utrecht: LOT.
2006 "Questions in legal monologues: Fictive interaction as argumentative strategy in a murder trial", *Text & Talk* 26 (3), 383-402.
Sala, Michele
2008 "Cultural and professional identities in international law", *Conference* Proceedings: Identity and Culture in English Domain-specific Discourse:
Methodological Issues and Preliminary Studies. University of Naples (Italy), October 2006.
2010 "Interrogative forms as professional identity markers in legal research articles". In: G. Garzone et al. (eds.) *Discourse, Identities and Roles in Specialized Communication*. Bern: Peter Lang, 301-320.
Scott, Mike
2007 *WordSmith Tools Version 5.0*. Oxford: Oxford University Press.
Solan, Lawrence
1993 *The Language of Judges*. Chicago: The University of Chicago Press.
Tiersma, P.M.
1999 *Legal Language.* Chicago: University of Chicago Press.
Tognini Bonelli, Elena – Manca, Elena
2002 "Welcoming children, pets and guest. A problem of non-equivalence in languages of 'agriturismi' and 'farmhouse holidays'". In: P. Evangelisti et al. (eds.) *English in Academic and Professional Settings: Techniques of Description / Pedagogical Applications.* 15 (2), 317-334.
Walter, Bettyruth
1988 *The Jury Summation as Speech Genre: An Ethnographic Study of What It Means to those Who Use It*. Amsterdam: John Benjamins.
Williams, Christopher
2005 *Tradition and Change in Legal English Verbal Constructions in Prescriptive Texts*. Bern: Peter Lang.
Ziegeler, D. P.
2006 "Mood and modality in grammar". In: Keith Brown et al. (eds.) *Encyclopedia of Language & Linguistics* (2nd edn.). Amsterdam: Elsevier, 259-267.

APPENDIX

EU1:	Case No. COMP/M.2788
EU2:	Case No. COMP/M.4123
EU3:	COUNCIL DECISION 2004/265/EC
EU3a:	MEMORANDUM of Understanding between the European Community and the National Tourism Administration of the People's Republic of China
EU4:	COMMISSION DECISION 2004/883/EC
EU5:	COMMISSION DECISION 1999/34/EC
EU5a:	ANNEX I to COMMISSION DECISION 1999/34/EC
EU5b:	ANNEX II to COMMISSION DECISION 1999/34/EC
EU5c:	ANNEX III to COMMISSION DECISION 1999/34/EC
EU6:	COUNCIL DECISION 86/664/EEC
EU7:	Council Decision 2005/923/EC
EU8:	COUNCIL DECISION 2006/516/EC
EU9:	COUNCIL DIRECTIVE 90/314/EEC
EU10:	COUNCIL DIRECTIVE 95/57/EC
EU10a:	ANNEX to COUNCIL DIRECTIVE 95/57/EC
EU12:	COUNCIL RECOMMENDATION 86/666/EEC
EU13:	COUNCIL DIRECTIVE 96/57/EC
EU13a:	ANNEX to 96/57/EC
EU14:	COUNCIL RESOLUTION 84/C_115/01
EU15:	COUNCIL RESOLUTION 86/C_340/01
EU16:	COUNCIL RESOLUTION 96/C_375/02
EU17:	COUNCIL RESOLUTION 96/C_155/01
EU18:	DIRECTIVE 94/47/EC
EU19:	PROTOCOL on the implementation of DIRECTIVE 94/47/EC
EU19:	PROTOCOL ON THE NEW MEMBER STATES JOINT DECLARATION ON IMPLEMENTATION ARRANGEMENTS.
EU19a:	ANNEX TO PROTOCOL ON THE NEW MEMBER STATES
EU20:	SPECIAL REPORT No 3/96 (97/C_17/01)

Inferentials: Fixed or not?*

Andreea S. Calude and Gerald Delahunty
University of Reading and Colorado State University

ABSTRACT

In this paper, we present arguments for analysing inferentials (which we class as a subtype of *it*-clefts) as partially formulaic. By exploring excerpts of spoken (New Zealand) English from the Wellington Corpus of Spoken New Zealand English, we establish that inferentials have formulaic tendencies: they are lexically limited, situationally bound, and relatively frequent (compared to other lexical bundles), and they serve a specific discourse function. However, they are not (perhaps, yet) fully established "fixed formulae" since they are semantically transparent, compositional, and non-fluent.

1. What are inferentials?

This paper is about the sentence types exemplified in bold:

(1) **It's not that I'm so smart. It's just that I stay with problems longer**. (Albert Einstein)

(2) Jake, come here buddy. Sit down. Look, **it's not that I don't care what you want. It's just that you're a kid, and what you want doesn't matter**. (*Two and a Half Men*, American sitcom)

(3) **It's not that I'm afraid to die**. I just don't want to be there when it happens. (Woody Allen)

* We are grateful to the participants in the Perpignan 2010 conference on "Fixed Phrases in English" for their insightful comments on an earlier draft of this paper. Any errors that remain are ours alone.

Various theories have been put forward about the analysis of these sentence types. A common view is that they are a type of cleft (Declerck 1992; Delahunty 2001; Koops 2007). Heggie (1998) sees them as copular constructions, but not of the cleft variety. Collins (1991) and Schmid (2009) propose that they are extrapositives. Finally, Pusch (2006, forthcoming) and Fraser (1999) refer to them as discourse markers.

We have argued elsewhere (Calude – Delahunty 2011) in favour of a cleft analysis for the inferential construction. The full debate and precise arguments which we take to support the cleft view are beyond the scope of this paper. Here, we briefly summarize our understanding of how inferentials are structured because their elements are relevant to the discussion of whether or not they are fixed.

2. A cleft analysis of inferentials

Cleft constructions privilege a clause constituent by placing it in focus position with the help of a copula and a few other elements that depend on the specific cleft type. The archetypal cleft is the *it*-cleft, but in languages rich in cleft types, such as English, there can be many others, e.g., *wh*-clefts, reversed *wh*-clefts, demonstrative clefts, *since*-clefts, *all*-clefts, and so on. The *it*-cleft is exemplified in (4).

(4a) It is **depression** that I fear most.

(4b) It is **the cookie jar** that I fear most.

(4c) It is **running out of gas** that I fear most.

Examples (4a) through (4c) show that *it*-clefts can focus noun phrases of various levels of complexity: from simple phrases as in (4a) to more complex ones as in (4c).

Noun phrases are not the only types of phrases which can be focused in *it*-clefts. As shown in (5) and (6) respectively, prepositional phrases and adverb phrases can also be focused.

(5) It is **in December** that we hope to go over to Spain.

(6) It is **intelligently, not speedily** that we want our employees to work.

Though they allow variants that we discuss below, *it*-clefts basically consist of an expletive *it*, a form of the copula, a focused expression, and a clause containing a gap of the same syntactic type as the focus. We will refer to the *it* + copula part as "the matrix," the post-copular constituent as "the focus," and the gapped clause as "the clause."

Sentences (1)-(3) and (7) show that a finite clause may occur in the focus position, and it is for this and other reasons that we view inferentials as a subtype of *it*-cleft.

(7) It is **that inferentials are too difficult** that I fear most.

However, many inferentials (particularly those in speech, see Koops 2007), occur without the gapped clause, as in (8).

(8) It is **that inferentials are too difficult.**

In itself, this is not problematic because other *it*-clefts exhibit a similar pattern, discussed by Hedberg (2000), Declerck (1988) and Huddleston – Pullum (2002), under the label "truncated *it*-cleft." Hedberg's classic example is reproduced below.

(9) Who ate the last cookie? It wasn't **me**.

Examples like (9) are *it*-clefts because the ellipted clause is generally recoverable from context, as in (10).

(10) It wasn't me **that ate the last cookie**.

This reasoning applies to the truncated inferential in (8): the ellipted clause may be recovered from context, as in (11).

(11) It is that inferentials are too difficult **that most concerns us.**

However, it is not always possible to recover the missing clause of a truncated inferential, e.g., it is not at all clear what the ellipted clause of the inferential bolded in (12) might be. (This is from the Wellington Corpus of Spoken New Zealand English, henceforth WSC. See section 3 for details, Holmes et al. (1998) for discussion, and the Appendix for a list of annotations used in the excerpts.)

(12) WSC DPC290

FR: [tells a story] … like i was very you know when <laughs> when we were kids we were always taught you gotta lock the car before you leave so i locked all the doors and <"> everything and so the um yeah so i locked his keys in there and so i told him he needed a spare key in the place sec spare set but he didn't you know and <quickly> then the other night </quickly> i know what happened to me

MQ: mm yeah yeah <laughs> he does **it's just he doesn't trust you** that's all cos he knows that you'll get a hold of it and <latch>

FR: <laughs> oh yeah

MQ: take it for a burn

FR: you reckon

Even though an appropriate clause may not be recoverable, we are justified in categorizing inferentials as *it*-clefts because of their lexical and syntactic commonalities and the fact that *it*-clefts and inferentials function very similarly in context, as we show below.

3. Data

The data for this paper consist of 55 inferentials taken from excerpts in the Wellington Corpus of Spoken New Zealand English (WSC), which contains approximately 250,000 words of spontaneous conversation. New Zealand English is probably the newest variety of English; its origins date back to the 19th century flow of Australian and British immigration. The most distinctive features of the New Zealand English variety pertain to its phonology (centralised vowels and various mergers), lexis (loanwords from the indigenous te reo Māori) and the widespread use of the final pragmatic particle *eh*. As far as grammar is concerned, New Zealand English does not have any completely unique variants, but rather exhibits a unique combination of features found in other varieties (mostly British English and its non-standard varieties), e.g., the use of *youse/yous* for the second person plural, the use of *she* in inanimate contexts (*she'll be right*), unmarked plurals (*that'll be fifty cent*), double comparisons (*more better*), demonstrative *them* (*them things, them people*), no clear distinction between *shall* and *will*, variability in past participle forms such as *proved* and *proven*, confusion between *bought* and *brought*, and so on (see Bauer 2007 for a summary). As far as we

are aware, no variation has been documented with regard to inferentials specific to New Zealand English.

Our WSC data were identified manually: because the constituent elements of inferentials occur in a broad range of constructions it was impossible to automate the search. Any corpus search for inferentials quickly turns up expressions that share many elements with the canonical inferential and thus raise the question of what to include in the data set. We found instances of expressions that are expectable grammatical variants of the canonical inferential: instances in which the copula is in the past tense; instances from which the conjunction *that* was missing; instances in which the clause was modified by adverbs such as *just*, *only*, *simply*, *actually*, and *not*; as well as instances with a modal, *could* or *may*, in the matrix. Because these are variants of the basic inferential licensed by the grammar and because they function contextually like inferentials but with modulations predictable from these grammatical modifications, we include them in our count.

We also discovered examples that seemed somewhat more distant from the canonical inferential, viz., those with the conjunctions *as if* and *like* in place of *that*. Because their meanings and discourse properties are similar to those of canonical inferentials and their differences from canonical inferentials are predictable from the meanings of *as if* and *like* we include these in our count also. (Calude and Delahunty 2011 provide an expanded discussion of this issue.) Table 1 lays out the variants.

Table 1. Inferentials in spoken New Zealand English

Types		Example	Spoken NZ English (WSC corpus)
Modified			
	negation	*it's not that/as if*	9
	just	*it's just that*	39
	epistemic modals	*it could/may be that*	1
	discourse markers	*it is well/you know/ i mean that*	2
Unmodified			
	plain unmodified	*it is that …*	4
TOTALS			55

4. The interpretations of inferentials

Positive cleft sentences are widely regarded as foregrounding the focus expression relative to some background, which, in the case of full canonical clefts, is the clause: the clause represents an open proposition and the focus represents the value of its variable (Delahunty 1984). In the case of truncated clefts, the focus is foregrounded relative to some contextually recoverable open proposition. Positive clefts assert and negative clefts deny that the focus is the relevant value for the open proposition. Canonical inferentials function similarly: the positive ones assert the relevance of the clause against a pragmatically determined local context; negative ones deny its relevance against such a context. (See Delahunty 2001 and Calude – Delahunty 2011 for more developed discussions of these claims.) We can see these interpretations in the following extracts.

4.1 That-*inferentials*

We assume that inferentials with *that* as introducer of the focal clause are the basic type of inferential and that inferentials with matrix negation or without a clause introducer are variants of this type. Delahunty (2001) is concerned solely with *that*-inferentials and argues that these inferentials are to be interpreted either as an aspect of the context in which some other expression is to be interpreted, or as an interpretation or reformulation of an immediately prior piece of text (an interpretation that does not seem relevant to any of our current examples). It is also argued there that the contextualized interpretations of *that*-inferentials are derivable from the interaction between the lexico-grammatical characteristics of the inferential, the principles of relevance theory, and the local context, with no need to stipulate any construction-specific procedural or idiomatic interpretation. The proposition represented by the inferential clause may function as either a premise or a conclusion relative to another contextually determined proposition, though the former seems to be more frequent, both in English and cross-linguistically (Delahunty – Gatzkiewicz 2000; Delahunty 2001). Inferentials interpreted as premises may be more richly interpreted as explanations, causes, reasons, and the like; inferentials interpreted as conclusions may be more richly interpreted as results or consequences (Delahunty 2001: 536-537). As an illustration, consider LL's negative *that*-inferential in (13):

(13) WSC DPC007

RR: you so you don't want to have to look over your shoulder to make sure you don't

→ LL: oh **it's not that i don't want to have to look over my shoulder** NOTHING should HAPPEN should OCCUR during those procedures ANY PART of it that's ALL formal the WHOLE lot right from the time the people come onto the marae (Maori: "meeting house") until the time everyone's LEFT the marae all right

RR's comment that LL would not want to look over his/her shoulder is prologue to what appears to be a result clause, *to make sure …*, and so functions as a premise. LL recasts this proposition as a negative inferential, thereby denying its relevance. The proposition represented by the inferential functions as a conclusion which follows from the remainder of his/her utterance: s/he won't want to look over her shoulder because NOTHING should HAPPEN, etc.

Because the basic inferential can be interpreted without resorting to any construction-specific semantic or pragmatic stipulations, we believe that the other inferential types may also be so interpreted, and that their interpretational differences from the basic type are due entirely to their lexico-grammatical differences from that type. In the following sections, we aim to show that this assumption is well-founded.

4.2 *Negative inferentials*

Negative inferentials work as their linguistic characteristics would predict – the inferential form triggers the interpretation that the focused clause is to be interpreted as a contextually marked assumption whose relevance is denied by the speaker.

In (14), KT describes the Māori classes she chose and her reasons for choosing them. Her initial utterance potentially implicates that she did not remain in the bilingual class beyond her first year, an interpretation consistent with her inferential. KT's inferential functions as a premise, which moots and rejects the proposition that she wanted to leave the bilingual class and go to the main stream as an explanation for her not continuing in the bilingual class. The sentence following the inferential (underlined) explicitly expresses her actual reason for leaving the bilingual class.

(14) WSC DPC240

KT: you know like i was in the bilingual class in my first year you know cos she was she we were just having a chat and um she said have you got a piece of maori in you are you part maori and i said yeah my dad's just under half and um <,> she said oh yeah you know there was a few teachers that were wondering about that some of the parents and stuff and <,> and i said yeah i was in the bilingual class in my first year **it wasn't actually that i wanted to leave the bilingual class to go to main stream in my second year** it was because i wanted to take a an advanced maori paper at <,> the varsity because i wanted to you know nurture <drawls> my language

4.3 Just-*inferentials*

Just-inferentials are the most frequent type in our data, and we assume that those without *that* are elliptical versions of *that*-inferentials modified by *just*. In (15), the speakers are discussing the possibility of increasing orders for their business, which they run from home.

(15) WSC DPC293

MK: okay <next utterance directed to person with tape recorder> okay just pause it <"> can you handle like two kits

→ FY: oh forgot about that i suppose we could **it's just i ca i haven't seen the books** so thanks to your sister's fantastic way of cleaning her room

MK: because

SS: <laughs>

MK: because er <latch>

FY: suppose we could

MK: yeah because i think you know with the er <,,> with <drawls> er kate

FY: kate and carmen and see the thing is there's n if carmen and mike can pay for it this week

FY's *i suppose we could* is a dispreferred response to MK's request *can you handle like two kits*; compare it to more positive alternatives such as *Sure*. FY neither accedes nor rejects MK's request, ostensibly because FY is not sure that they can handle two kits. The inferential presents the proposition

i haven't seen the books as a contextual premise from which it would follow that FY would not know whether they could handle two kits or not.

According to Nevalainen (1997), *just* is a focusing particle meaning 'merely' and 'only', and 'exactly P' where "P" is not of great importance. (See also Quirk et al. 1985: 604.) According to Aijmer (2002: 158), the core function of *just* as 'exactly' and 'only' is a procedural marker indicating an "indexical relation to the speaker's attitudes or emotion towards a discourse event", so *just* always carries evaluative overtones. Consequently, the inferential in (15) indicates that FY's not having seen the books is the only reason why they might not be able to handle two kits, and it also suggests that this reason is of no great importance, and that MK can reasonably expect a positive response once FY has seen the books.

We note that because the clause of an inferential is interpreted as special in its context, it can be used to counter contextually possible assumptions or interpretations, which may arise from the prior discourse, as in (15). (The segment understood as the inferential trigger is underlined.)

4.4 (Not) as if-*inferentials*

Our hypothesis regarding the interpretation of inferentials with *like*, *not like*, *as if*, and *not as if* is that their interpretation is the same as that of *that*-inferentials except in so far as (*not*) *like/as if* differ from *that*. We begin with a discussion of *(not) as if* inferentials and then deal with the *(not) like* variant.

There is relatively little research on either of these inferentials, though Huddleston and Pullum (2002) include some suggestive remarks. According to Huddleston – Pullum (2002: 1146), *as* denotes comparison, and *if* "is primarily conditional" and thus has a "close relation" with *though*, "which is primarily concessive" (2002: 737). They claim that *as if* may function as a "single compound preposition," which, we believe, denotes a sense of hypothetical comparison (2002: 1151). This sense may be quite "attenuated" (2002: 1151) in certain contexts, and consequently, in some instances, *as if* may be replaced by *that* or its zero alternant without change of meaning, and so may be interpreted as merely suggesting the truth of a proposition rather than (strongly) asserting it. This is especially the case after *appear, feel, seem, sound,* and *be*, which may induce a "medium strength epistemic modality."

Biber et al. (1999: 840-841) claim that with non-finite clauses, *as if* and *as though* indicate that the "adverbial clause is showing similarity but is

not to be taken factually." This is consistent with the analysis we developed above.

We found only two *as if* inferentials in WSC and both were negative so we use a positive *as if* inferential from a written corpus of New Zealand English to begin our discussion of *(not) as if* inferentials.

(16) WWC SECTION F, F42 186-194
The Sunday News used to be the main proponent of the idea of celebrity in New Zealand. It was in that tabloid beloved of life's losers that we first read about Graeme Thorne's perm and much other such trivia. **It was as if successive editors had a list of so-called personalities from which they never really deviated.** It is probably still pasted up in the news-room, slowly yellowing under the harsh fluorescent lights. My guess is that it includes the old names Ray, Bob, Max, Marilyn, the other Ray and Howard. You should know the surnames. They've been around for years.

This inferential can be interpreted as a hypothetical premise from which the celebrity of Graeme et al. would follow, viz., the editors acted as if they drew the names of the celebrities from a list from which they never deviated. Thus this *as if* inferential functions as we predicted.

The inferential in (17), from WSC, shows that *not as if* inferentials also function as we predict.

(17) WSC DPC032
AW: well there's only there's only five or six in the race **it's not as if they're racing going up three wide ra round fields of eighteen** they're only going round fields of six they race sort of there'll be one in the front one <long pause> on the e one on th the trail and one on the outer behind the horse on the on the trail not facing the breeze

In the conversation from which (17) is taken, two people are discussing horse racing and specifically the differences between real racing and practicing. AW's description contrasts *going round fields of six* [horses] with *going round in fields of eighteen* [horses], and rejects the conclusion that in the former the horses are racing. This interpretation is supported in the utterance immediately following the inferential when AW characterizes going round fields of six as only "sort of'' racing. This is consistent with Huddleston

and Pullum's remark that "*It's not as if he wasn't trying* … is used to deny a proposition that might otherwise have been deduced (perhaps he didn't perform as well as expected)" (2002: 1152, fn. 36).

4.5 (Not) like-*inferentials*

Even though our negative occurrences of *like* inferentials outnumber our positive ones, we begin here also with the latter as it is more basic than the former:

(18) WSC DPC326

JI: and she was the legal advisor for er ronnie burch

AL: right <latch>

JI: you know <drawls> when when he was yeah race relations yeah <with creaky voice> mm but um

AL: yeah race relations oh <drawls> good so she didn't have a problem getting a job i suppose when her when er <unclear word>

→ JI: no no but sh she found that particular job very stressful **it's like she doesn't think she'd like to go back into it** you know cos she was always dealing with problems

AL: <drawls> mm right

Like, whose basic meaning reflects similarity, is being grammaticalized as a marker of reported speech and thought (Romaine and Lange 1991). Our analysis is consistent with this, but we suggest that by virtue of its basic meaning and its grammaticalization, *like* in inferentials is a marker of "interpretive use'' (Sperber and Wilson 1995: 224-231). That is, the proposition represented by the clause introduced by *like* is to be interpreted as, to one degree or another, resembling a proposition from which relevant contextual effects would follow. Crucially, the proposition represented by the inferential clause is not a proposition assumed by the speaker, but it merely resembles some such proposition. In (18), the proposition represented by the inferential clause is presented as similar to a proposition which is to function as a conclusion that would follow from the proposition represented by the immediately prior sentence, *she found that particular job very stressful*. This cause and effect relationship is made explicit by the conjunction "cos" which introduces the sentence that follows the inferential, *she was always dealing with problems*.

Not like inferentials reject the relevance of the proposition represented by the inferential clause as a more or less faithful interpretation of the proposition entertained by the speaker. Thus the negative inferential in (19) rejects the potential characterization of the situation as "they NEED someone". Because it is presented in inferential form, this proposition functions as a contextual proposition, in this context, most likely as an explanation, that is, as a proposition from which *They wouldn't want Thomas* would follow. This is consistent with the positive inferential that follows the *not like* one, *It's just they're just doing it as a favor because Susannah's a mate*. We interpret this as a premise from which it would follow that the Wilkins would have AC come *over there … every week*.'

(19) WSC DPC059

AC: well missus wilkins said i could do it over there i mean every week but they they wanted me to do it every week
BS: yeah pity in some ways isn't it because it's quite good money
AC: mm
BS: do you think thomas would do it
AC: they wouldn't they wouldn't want thomas
BS: mm
→ AC: **it's not like they NEED someone** it's just they're just doing it as a favour because susannah's a mate
BS: <drawls> yeah <">
AC: not like they need anyone
BS: i've OCCASIONALLY thought that you could actually do some work for kelvin and sharon but i'm not sure
AC: but they i should just do it for them for free

It is our intuition that in the examples above, *(not) like* may replace *(not) as if*, with only a stylistic shift; we find the *(not) as if* variants to be somewhat more formal than the *(not) like* variants. For example, (16) is from the written portion of the Wellington Corpus of New Zealand English and so we must assume that the author (and perhaps editors) chose their words carefully when they wrote *as if* instead of *like*. But given that substituting *like* for *as if* seems to have no effect on the interpretation of the text, we might assume that the choice merely reflects a different stylistic level. We believe that *like* indexes an informal context because of its origins in casual conversational use by teenagers (though they are not the sole users, see Miller 2009). As a result we find this use of *like* in informal, unplanned discourse and in

representations of such discourse, for example, in fictional renditions of speech. *As if*, on the other hand, cues a higher stylistic level, so we find it in more formal contexts, such as academic prose.

5. What is "fixed" and how might we identify it?

The notion of fixed or formulaic language has come to be used in opposition to that of novel or creative language, that is, expressions which are generated by the syntax and lexis of a language.

In contrast, fixed phrases are not produced with the aid of these rules, they are recurring expressions which are thought to be stored whole in memory, behaving much like single words. Because fixed phrases have come under scrutiny in various fields of linguistics and psycholinguistics, the phenomena appear in the literature under a variety of different labels (47 according to Wray 2002), some of which are given below:

frozen phrases	*idioms*
formulaic expressions/language	*collocations*
fixed expressions	*ready-made chunks*
lexical bundles/phrases	*composites*
epistemic phrases	*recurring utterances*
prefabricated patterns	*conversational routines*

As summarised by Edmonds (2010), two major trajectories have been explored in the understanding of fixed phrases. One of these is a functionally and pragmatically motivated path which focuses on the recurring and fixed nature of these expressions, encompassing work by Altenberg (1998), Bardovi-Harlig (2009, 2010), Coulmas (1979), Kuiper et al. (2007), Pawley (2008), and others (see Edmonds 2010: 14). Within the functional-pragmatic approach, a fixed phrase is an invariable or minimally variable expression which conventionally uses a particular turn of phrase in a specific situation, community, or culture (or combination of these). For example, *fish and chips* refers to a specific type of meal, where the potatoes are cut into strips of a given size (which may vary across countries) and then deep-fried, and the fish is usually a piece of cod which is battered and served (ideally) wrapped in a newspaper sheet, different and distinct from *chips and fish* (which could be any kind of chips together with any kind of fish cooked in any way).

A further example is the use of the phrase *to be someone's shout* which in New Zealand English refers to someone's turn to pay for the following round of (typically) drinks. This phrase is used conventionally in this way only by members in (or familiar with) the New Zealand community, and standing up on the way to the bar in a British pub uttering the phrase *it's my shout* would cause some confusion.

A second strand of research concerning fixed phrases comes from psycholinguistics, developed in work by Biber – Conrad – Cortes (2004), Weinert (1995), Wray (2002), Wray – Perkins (2000), and others. These studies privilege the stored aspect of fixed phrases, with the main characteristic being the claim that they are retrieved whole from memory and thus behave like single words. Prime examples include *by and large*, whose internal grammar is frozen and no longer in line with the grammar of modern English, sayings such as *nice guys finish last* and *life is like a box of chocolates*, or greetings and endings like *how do you do*, *may I help you*, and *yours sincerely*.

The large and growing level of interest in fixed phrases is reflected in the existence of the *Yearbook of Phraseology* series. Despite the vast body of work to date (only some of the works have been mentioned in the brief summary above), fixed phrases remain slippery and elusive in that there is arguably no widely-agreed upon definition or term (though the label *formula* is perhaps gaining the most support), or a definitive set of criteria for identifying these recurring expressions.

In spite of the lack of a consensus, some criteria have been proposed as generally characteristic of fixed phrases and are thus useful in identifying them. Like many other notions in linguistics, fixedness is not a binary feature, but rather encompasses a continuum between something which is more or less formulaic (or more or less novel, depending on how one looks at it). Hence these criteria are not a necessary and sufficient set, but rather a set of heuristics which help identify those expressions that occupy the more fixed end of the continuum. The criteria are listed below in (A) – (K), and are based on the detailed summary of the literature presented in Edmonds (2010: 19-31):

(A) Multiword/multimorpheme
(B) Invariability
(C) Frequency
(D) Community-wide use
(E) Situational boundedness
(F) Syntactic coherence

(G) Semantic opacity
(H) Noncompositionality
(I) Discourse planning and greater fluency
(J) Complexity
(K) Overextension

Criterion (A) has to do with whether or not a single word may count as a fixed expression. Opinions diverge on this issue. The debate is not relevant to the present study, as the inferential matrix is a multiword expression anyway. Criteria (J) and (K) relate specifically to learner uses. And because the data analysed here are exclusively produced by adult native speakers, these two properties cannot be investigated in this study, and will not be discussed further. Therefore, the relevant criteria to our work on inferentials are those given in bold, from (B)-(I).

Table 2 below lists these eight criteria, exemplifying what an idealised fixed phrase might be like, relative to each of these properties.

Table 2. Fixed phrase criteria relevant to the current study

Criterion	If expression X is fixed then…	Example
1	2	3
Structural and/or lexical invariability	X is largely invariant, though some open slots are permitted	*What's this A doing in my B*, where A and B can refer to any NP which fulfills the pragmatic requirements
Frequency	X occurs with relatively higher frequency than expected (e.g., more frequently than it could have occurred given the topic, context, etc. at hand)	*I don't know what* (Biber et al. 1999: 996)
Community-wide use	X is known to an entire speech community, not just to restricted sets of speakers	On the other hand
Situational boundedness	X is associated with a particular situational context, such as a social or pragmatic situation	*How do you do*? is associated with greetings

1	2	3
Syntactic coherence	X does not cross constituent boundaries and applies to full phrases, NPs, VPs, PPs, full clauses, etc.	All things considered
Semantic opacity	The meaning of X is not clear from its component parts	*Kick the bucket* has two meanings, one of which (the idiomatic one) is not transparent from its component words
Noncompositionality	The grammar of X is frozen and may be at odds with current patterns of the grammar of the language it belongs to	*By and large* is no longer in line with grammatical patterns of modern English
Discourse planning and greater fluency	Because X is fixed, it functions much like a single word, it is uttered more quickly than other phrases might be, it is not interrupted by pauses, discourse markers, false starts, etc.	*How do you do* is typically uttered without any pauses, discourse markers or interruptions, and faster than other four word phrases

The first observation to be made is that some of the criteria in Table 2 are not easily operationalisable. In particular, the frequency criterion is notoriously difficult to apply as it is rare that anyone can accurately ascertain just how many times a given phrase could have occurred in a text. In practice, this criterion often becomes a question of collocationality instead (e.g., *kick* and *bucket* more often occur together than not).

Second, these criteria cannot be used in a vacuum, but must be applied to an expression whose fixedness we want to ascertain within a given corpus or data set. We cannot know whether a particular expression has greater fluency or whether it is situationally bound unless we examine how it is actually used. The full range of possible uses of an expression may not always be fully borne out in actual use. Complementarily, real linguistic documentation may surprise us in that the use of a given expression may be wider than previously realised.

Third, it must be remembered that few phrases will exhibit all of these criteria. The majority of expressions with fixed tendencies will exhibit some of these properties but not others. As might be expected, the more of these criteria an expression meets, the more certain we can be that it is fixed, and the more fixed it is likely to be.

6. Are inferentials fixed?

As mentioned in the introduction, we conducted an investigation of inferentials such as those exemplified in (1)-(3) with the aim of establishing their best syntactic analysis. Our findings, discussed in detail in Calude – Delahunty (2011), suggest that the inferential is a type of cleft. However, while exploring the properties of inferentials in WSC, we developed the hypothesis that inferentials might be fixed, at least in spoken English. (We cannot say anything about inferentials in written English.) The reasons for this hypothesis are threefold.

First, our attention was drawn to the fact that the majority of inferentials seemed to have a recurring, relatively invariant structure of the form [*it's (just) that* S]. Consider the following examples from WSC:

(20a) WSC DPC129

CH: hey by the way can i borrow a pair of your earrings <,> **well it's just that** <laughs> <latch>
RG: it depends <,,> which pair
CH: oh well which pair are which pair are you going to wear and i'll borrow another <latch>
RG: the new ones i just bought today
CH: oh well then how about i wear those big ones with the bl with the crystals

(20b) WSC DPC136

AT: why mum was complaining about how GRUESOME the murder was
BD: oh yeah and then they exorcised the demon out of this guy who last episode that was quite fun
AT: i'm sure she loved that <laughs>
BD: and it was really weird um it <,> **it's just that they don't know what happened** but they actually DO know what happened all the sprinklers went on because um you don't know any of the characters <laughs> so it's really hard to explain this to you but but lucy was having this conference with andy and dick who are the possible fathers of her child
AT: sounds like molly dodd <laughs>

(20c) WSC DPC138

AA: oh that's all he said cos it at first i was just saying about how i jim's about the <,,> only one i've had problems with

BC: <drawls> oh right yeah
AA: cos like jim's the only one <,> that won't always <latch>
BC: yeah i know he's just harder to get on with
AA: yeah oh **it's just that i don't think that jim does it on purpose** i think it it's him <latch>
BC: it's just him oh yeah
AA: yeah it's his personality <latch>

Second, similar constructions have been argued to be formulaic (particularly in spoken data). A study of demonstrative clefts in the same corpus argued that this cleft type is fixed, encompassing a formula with a few open but predictable slots, and a specific discourse function (see Calude 2009a, 2009b). The clauses in bold in (21a) and (21b) are typical examples:

(21a) WSC, DPC096 (from Calude 2009a, ex. 9, p. 65)
BG: oh no she is lovely she's gossipy though
AT: mm
BG: very gossipy like bill **that's where Bill get's it from**
AT: <unclear word> oh he is a little gossip talking about Mike Furley

(21b) WSC, DPC214 (from Calude 2009a, ex. 18, p. 69)
BH: the brace helps to hold you upright <,,,>
UV: the only thing for a sore back is bed rest
BH: well **that's what they say** eh
UV: yup
BH: and heat

Similarly, Hopper and Thompson (2008) show that English *wh*-clefts and extraposed clauses as well as German *wenn*-clauses are also fixed. These complex clauses behave more like monoclausal units than like biclausal complexes and should thus be analysed as "single, partly formulaic clauses deployed by speakers in managing interactional discourse" (2008: 99). For example, the pseudoclefts in (22a) and (22b) start off a set of instructions or explanations, and have an invariant recurring pattern:

(22a) (from Hopper – Thompson 2008, ex. 9, p. 6)
So then what you do is, you sprinkle the fifth-graders out evenly.

(22b) (from Hopper – Thompson 2008, ex. 13, p. 7)
What we do, then that's … that's where the ferrier comes in.

Note that the recurrence of *you* in (22a) would render it ungrammatical in a highly edited text, and that there is no grammatical connection between the introductory *wh*-expression and the clause it prefaces, contrary to what would be expected under a biclausal analysis. These types of clefts were also noted in the Map-Task Dialogue corpus of spoken Scottish English analysed by Miller – Weinert (2009), and are so widespread in spoken language that they have now entered textbooks of English grammar (see Miller 2000 and 2011).

Third, as the body of work just mentioned reiterates, spoken language in general (and spontaneous spoken language in particular) appears to make systematic use of formulaicity and the recycling of structures and phrases. In this medium, many constructions are simplified and invariant, consisting of a set of predictable patterns associated with specific discourse-related properties and interactional characteristics. The conditions under which spoken language is produced and parsed, and the functions for which it is used have a substantial effect on its character. Large scale analyses (see for example Miller – Weinert 2009 and Biber et al. 1999) document a reduction in the variety of forms and structural integratedness in spoken language in comparison with written language. The most affected expressions are complex constructions, particularly those involving subordination. These are 'reduced' to only a few possible patterns, in part due to the decreased cognitive loading required for their encoding and parsing, and in part, to their acquiring specialised interactional functions.

Given these observations, we might expect our analysis of the inferential cleft in spontaneous spoken conversation excerpts from the WSC to reveal that it is at least partially formulaic, thus categorizing it with pseudo-clefts, demonstrative clefts, extraposed clauses, and German *wenn*-clauses. We test our hypothesis using the criteria we listed above.

As we noted, the great majority of inferentials can be partially described by the formula [*it's (just) that S*]. We say "partially" here because such linear expressions omit whatever hierarchical organization might be present, and the specification that *it* is expletive. However, it is noteworthy that *it* and *is* are contracted in every example in our data where they may be, which we take as support for our hypothesis.

This formula may be adapted to include the remaining inferential types:

(23) It (modal) (not) BE (adv) (conjunction) S

implying that expletive *it*, the copula, and the complement clause are necessary and sufficient elements. That we can describe them with formulae of this sort suggests that inferentials may be fixed, though we must add that

it and *be* are the minimum required to focus an expression, which is the purpose of the sentence type.

Frequency. Our data show that inferentials occur 55 times in approximately 250,000 words, which is approximately 220 times per million words of conversational English. We also know that they occur in other registers, though we do not know how frequently. Biber et al. (1999) identify a lexical bundle as "a recurring sequence of three or more words" and that four word bundles must occur at least ten times per million words, though longer bundles occur less frequently, and must occur in at least five different texts. As inferentials occur far more frequently than Biber et al's. criterial frequency, we might conclude that they are at least as fixed as lexical bundles. However, Wray (2002) cautions against relying on frequency as a fixity criterion.

Community-wide use. Our data show that inferentials occur in spoken New Zealand English and Koops (2007) shows that they occur in conversational US English, so we can reasonably conclude that they are incommunity-wide use. However, as the grammar and vocabulary of English are in community-wide use, we would have to conclude that any expression generatable by the grammar would be fixed, a patently absurd conclusion.

Situational boundedness. An expression is situationally bound if it is consistently used for a particular social or pragmatic purpose. Inferentials perform a single pragmatic function, viz., asserting or denying the special relevance of the proposition represented by the clause. When we combine this criterion with the relative invariability of inferentials, we conclude that they are partially fixed.

Syntactic coherence. Inferentials are generatable by the grammar and lexis of English and so are syntactically coherent.

Semantic opaqueness. The semantics of inferentials is a function of their lexis and syntax and is therefore transparent.

Noncompositionality. Because the semantics of inferentials is transparent, it is compositional. This criterion, the opacity criterion, and the syntactic coherence criterion together indicate that the inferential forms are generatable by the grammar and lexis of English and that their meanings are a compositional function of those forms, and therefore indicate that inferentials need not be fixed.

Fluency. We have found several instances in which the inferential matrix is interrupted by fillers, e.g., *it's um that i mean public transport …*; *it's just you know <,> it was like <leave me alone>*, suggesting that the matrix is generated or at least generatable analytically and so not stored and produced holistically as a single word-like unit.

Discourse planning. Because fixed expressions are likely to be relatively frequent and therefore to come readily to mind, we should reasonably expect them to be produced at points in discourse where speakers need space to plan what they are about to say. Given that, we should also expect that the utterance that follows the production of a fixed expression need not be grammatically integrated with it, as Hopper and Thompson (2008) discovered with English *wh*-clefts and extrapositives and German *wenn*-clauses. Because inferentials have a specific pragmatic or discourse function and are composed of elements that are frequent and easily processed, they should be readily exploitable in this way. However, inferential matrixes do not have the intonational contour Hopper and Thompson observed with the expression types they studied and which "projects" further talk by the producer, and we have so far not found examples that display the lack of connectedness between matrix and complement that Hopper and Thompson found. This too suggests that inferential matrixes may not be fixed.

How are we to interpret the fact that our application of these fixity criteria to inferentials gives us inconsistent results? Because inferentials are describable with a partially fixed formula with very limited possibilities in its variable slots, and they are in frequent, situationally bound, community-wide use, they appear to be formulaic. However, their matrixes are syntactically coherent, are semantically compositional and transparent, are motivated by the need to focus a clause, may be interrupted by fillers, and do not display either the kind of intonational separation of matrix from complement, intonational projection of the complement, or grammatical independence of the complement from the matrix that Hopper and Thompson identified in their target expressions, which we use as a benchmark for formulaicity, suggesting that they may be constructed as needed rather than stored and produced whole.

It seems to us that being describable as a formula, relative frequency, situational boundedness, and community-wide use do not entail fixity. Rather they are conditions that are consistent with fixity and which perhaps predispose expressions to fixity. We think that English inferentials, at this point in their history, are not fixed, though they may be on the cusp of becoming so, as the consistency with which *it* and the copula are contracted suggests. This development is also indicated by the fact that their matrixes are created out of the kinds of linguistic items – *it* and the copula – that would lend themselves to fixation and eventually to reduction and univerbation. That is, items which occur frequently, are readily accessed and processed, but are non-salient (Hudson 1998).

7. Conclusion

Spoken inferentials fail several of the tests for formulaicity: they are generatable by the current grammar of English; they are semantically transparent and compositional; their pragmatic and discourse effects are predictable from their forms and meanings; they are not always produced fluently and without interruptions. However, though predictable from their grammar, inferential matrixes are lexically very limited, allowing only a very few possibilities. All inferential variants have an expletive *it* subject, a form of *be*, and a complement, which are essential for their interpretation and function; they allow limited tense variability on the copula, the possibility of modals (though only two show up in our data), and a very restricted set of adverbs and clause introducers. They are relatively frequent and are situationally bound because, however we define or delimit them, they have a specific discourse purpose. To resolve these inconsistencies we suggest that they may be at the beginning stages of fixation.

REFERENCES

Sources

Bauer, Laurie
1993 *Manual of Information to Accompany the Wellington Corpus of Written New Zealand English*. Wellington: Department of Linguistics, Victoria University of Wellington. (http://khnt.aksis.uib.no/icame/manuals/wellman/INDEX.HTM).

Holmes, Janet, Bernadette Vine, and Gary Johnson
1998 *Guide to the Wellington Corpus of Spoken New Zealand English*. Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington. (http://khnt.aksis.uib.no/icame/manuals/wsc/INDEX.HTM).

Special studies

Aijmer, Karin
2002 *English Discourse Particles*: *Evidence from a Corpus*. (Studies in Corpus Linguistics 10.) Amsterdam: John Benjamins.

Altenberg, Bengt
1998 "On the phraseology of spoken English: The evidence of recurrent word-combinations". In: A. P. Cowie (ed.) *Phraseology: Theory, Analysis and Applications.* Oxford: Clarendon, 101-122.

Bardovi-Harlig, Kathleen
2009 "Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics", *Language Learning 59*, 755-795.
2010 "Recognition of formulaic sequences in L2 pragmatics". In: G. Kasper et al. (eds.) *Pragmatics and Language Learning*. Vol. 12. Honolulu: University of Hawai'i, National Foreign Language Resource Center, 139-160.
Bauer, Laurie
2007 "Some grammatical features of New Zealand English", *New Zealand English Journal 21*, 1-25.
Biber, Douglas et al.
1999 *Longman Grammar of Spoken and Written English*. Essex: Longman.
Biber, Douglas – Susan Conrad – Viviana Cortes
2004 "If you look at…: Lexical bundles in university teaching and textbooks", *Applied Linguistics 25*, 371-405.
Calude, Andreea S.
2009a "Formulaic tendencies of demonstrative clefts in spoken English". In: R. Corrigan et al. (eds.) *Formulaic Language*. Vol. 1: *Distribution and Historical Change*. Philadelphia: John Benjamins, 55-76.
2009b *Cleft Constructions in Spoken English*. Berlin: VDM Verlag.
Calude, Andreea S. – Gerald P. Delahunty
2011 "Inferentials in spoken English", *Pragmatics 21,* 307-340.
Collins, Peter
1991 *Cleft and Pseudo-cleft Constructions in English*. London: Routledge.
Coulmas, Florian
1979 "On the sociolinguistic relevance of routine formulae", *Journal of Pragmatics 3*, 239-266.
Declerck, Renaat
1988 *Studies on Copular Sentences, Clefts and Pseudo-clefts*. Leuven: Foris.
1992 "The inferential *it is that*-construction and its congeners", *Lingua 87*, 203-230.
Delahunty, Gerald P.
1984 "The analysis of English cleft sentences", *Linguistic Analysis 13*, 63-113.
2001 "Discourse functions of inferential sentences", *Linguistics 39*, 517-545.
Delahunty, Gerald – Laura Gatzkiewicz
2000 "On the Spanish inferential construction *ser que", Pragmatics 10*, 301-322.
Edmonds, Amanda
2010 *On the Representation of Conventional Expressions in L1-English and L2-French*. Ph.D. dissertation, Departments of French and Italian and Linguistics, Indiana University.
Fraser, Bruce
1999 "What are discourse markers?", *Journal of Pragmatics 31*, 931-952.
Hedberg, Nancy
2000 "The referential status of clefts", *Language 76*, 891–920.

Heggie, L.A.
1998 *The Syntax of Copular Structures*. Ph.D. dissertation, Department of Linguistics, University of Southern California.
Hopper, Paul – Sandra Thompson
2008 "Projectability and clause combining in interaction". In: Ritva Laury (ed.)
Crosslinguistic Studies of Clause Combining: The Multifunctionality of *Conjunctions*. Amsterdam: Benjamins, 99-123.
Huddleston, Rodney – Geoffrey Pullum
2002 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Hudson, Jean
1998 *Perspectives on Fixedness: Applied and Theoretical*. Lund: Lund University Press.
Koops, Christian
2007 "Constraints on inferential constructions". In: G. Radden et al. (eds.) *Aspects of Meaning Construction.* Amsterdam: John Benjamins, 207-224.
Kuiper, Koenraad et al.
2007 "Slipping on superlemmas: Multi-word lexical items in speech production", *The Mental Lexicon* 2, 313-357.
Miller, James
2009 "LIKE and other discourse markers". In: P. Peters et al. (eds.) *Comparative Studies in Australian and New Zealand English: Grammar and Beyond*. Amsterdam: Benjamins, 317-338.
2000 *An Introduction to English Syntax*. (Edinburgh Textbooks on English Language.) Edinburgh: Edinburgh University Press.
2011 *A Critical Introduction to Syntax*. (Continuum Critical Introductions to Linguistics.) London: Continuum.
Miller, James – Regina Weinert
2009 *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Oxford University Press.
Nevalainen, Terttu
1997 "Recycling inversion: the case of initial adverbs and negators in Early Modern English", *Studia Anglica Poznaniensia* 31, 203-214.
Pawley, Andrew
2008 "Developments in the study of formulaic language since 1970: A personal view". In: P. Skandera (ed.) *Phraseology and Culture in English*. Berlin: Mouton de Gruyter, 3-48.
Pusch, Claus
2006 "Marqueurs discursifs et subordination syntaxique: La construction inférentielle en français et dans d'autres langues romanes". In: M. Drescher – B. Frank-Job (eds.) *Les marqueurs discursifs dans les langues Romanes: Approches théoriques et méthodologiques*. Frankfurt: Peter Lang, 173-188.

In preparation "Pragmatic markers involving subordination in Romance: Do they structure discourse or comment on it?"
Quirk, Randolph et al.
1985 *A Comprehensive Grammar of the English Language*. London: Longman.
Romaine, Suzanne – Deborah Lange
1991 "The use of LIKE as a marker of reported speech and thought: A case of grammaticalization in progress", *American Speech* 66, 227-279.
Schmid, Hans-Jorg
2009 "Rare but contextually entrenched: The English *not-that* construction". Paper presented at the *International Conference on the Linguistics of Contemporary English*, University of London.
Sperber, Dan – Deirdre Wilson
1995 *Relevance: Communication and Cognition*. Oxford: Blackwell.
Weinert, Regina
1995 "The role of formulaic language in second language acquisition: A review", *Applied Linguistics 16*, 180-205.
Wray, Alison
2002 *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
Wray, Alison – Michael Perkins
2000 "The functions of formulaic language: An integrated model", *Language and Communication 20*, 1-28.

APPENDIX

WSC annotations

<drawls>	speaker drawls
<latch>	overlapping speech
<laughs>	speaker laughs
<„>	longer than 1 second pause
<,>	1 second pause
<reads> … </reads>	portion given between tags was read by the speaker
<quickly>	speech portion is uttered quickly
<quietly>	speech portion is uttered in a quiet voice
<softly>	speech portion is uttered softly
<unclear word>	speech is inaudible or incomprehensible
<with creaky voice>	speech portion is uttered with a creaky voice

The prepositions *at* and *to*: Opposite points of view

Patrice Larroque
Paul Valery University, Montpellier

ABSTRACT

The purpose of the present article is to describe the syntactic and semantic functions of the prepositions *at* and *to*. Starting from the false argument that *at* is associated with hostility, this paper tries to show that *at* and *to* are opposites of the same notion (i.e. have a common origin) and therefore their meanings depend on the position from which the action or state is viewed. *At* refers to the receiving end while *to* indicates orientation towards a goal seen from the origin of the act.

The hostility conveyed by *at* ultimately depends on the semantic realization of its complementation which implies reception and is related to stability (nouns or -ING forms). *To*, on the other hand, remains ambiguous and unstable as regards the aim to be achieved. As such *to* is apt to acquire a metalinguistic value and introduce a verb base.

1. Introduction

This paper follows a discussion I had a few years ago with a colleague who argued that the preposition *at*, unlike *to*, was connected with hostility in some way. I countered the argument by evoking the case of *smile at somebody* which he dismissed as an exception. From that point of view, there are many exceptions to which *hint at*, *gawp at*, *look at, gaze at*, and *wink at*, for example, belong. What he probably meant was that the prepositional object of *at* can be regarded as a target, and that hostility concentrates on that particular point, making the phrase extremely assertive (Lapaire – Rotgé 1991: 89). My impression, after delving into the subject, is that not only is the hostility conveyed by *at* one of the many semantic realizations of the prepositional phrase, but *at* and *to* must be paired as opposites of the same notion, denoting opposite viewpoints relative to a given landmark.

In fact, *at* and *to* have a common origin and can be used to express location and direction towards a place or an object, but they have developed various and seemingly opposite interpretations and meanings, ranging from stating where something is (*at*) to the generic notion of transition (*to*).

Before adding further comments on the hostile connotations of *at*-PPs (prepositional phrases headed by *at*), I suggest describing the syntactic and semantic functions that the prepositions *at* and *to* fulfill in speech, and showing how they can be linked to the position from which the verbal happening is viewed. To begin with, I will consider the etymology and linguistic definitions of both prepositions.

2. Etymology and definitions

According to Quirk – Greenbaum (1973: 143), "a preposition expresses a relation between two entities, one being that represented by the prepositional complement." Its function then is to introduce a subordinate element (Rapatel 2010: 11-12) and determine various kinds of relational meaning (Jespersen [1933] 2002: 69). In addition, Huddleston and Pullum (2002: 598) consider that prepositions are heads of phrases, which increases the set of words that are traditionally assigned to the category of prepositions, and allows them to take dependents other than noun phrases (e.g. *down by the riverside, until recently, for now, for less than $ 40*).

Let us return now to the morphemes *at* and *to* whose description we are considering. They both derive from the IE base **de/*do/*dō*. The original meaning of the IE base is that of "putting in contact" which *at* and *to* still have (Bourquin 1990: 116). *At* was lost in Southwest dialects (and the Germanic dialect) and replaced by *to*.[1]

At can be defined as a preposition and verbal prefix (cf. *ado, at +do*: trouble, dealings) denoting position and motion towards. It is used to indicate a point or place occupied, a location, an amount, an occupation, a state or condition (Webster's Dictionary). *To* is a preposition expressing motion or direction towards an object, addition, or the notion of the dative;[2] with infinitive meaning it bears the sense of "for the purpose of", or "with the object of (doing something)", hence serving without meaning as a sign of the infinitive. *To* can also be an adverb in expressions like *to and fro*, *pull*

1 In Scandinavian *to* was lost and replaced by *at* (Hoad 1980: 26).

2 Old English had a preposition *to* which took the dative: God cwæþ to þam wife, "*God said to the woman*" (cited in Freeborn 1998: 68).

the door to, or (*after) he came to*, and appears in compound *to-do* meaning a lot of excitement about something (e.g. *They made a great to-do over the dinner.*). It should be noted that *ado* and *to-do* are very close in meaning.

It is instructive at this point to stress the fact that both definitions exhibit direction/motion towards (an object or a contact point). They indicate positive position and direction in the sense that they express movement or position with respect to a destination as opposed to the negative character of *(away) from*, for instance, which marks the source location (Quirk – Greenbaum 1973: 147-148).

We now turn to examples and the descriptions of the prepositions and prepositional phrases.

3. Examples and descriptions

Prepositions were present in Old English, and they were associated with inflectional cases, most often the dative (*to, mid, on, of,* etc.). In a prepositional phrase the noun was usually in either the dative or the accusative case, according to the preposition (Quirk – Wrenn 1957: 68; Mitchell 1985: 497-498). Prepositions originally have concrete meanings: they express locational relations in time and space. But through frequent use, the verbalization of experience (Croft 2010), and the speaker's desire for expressiveness (Meillet 1912), a number of the most common prepositions have acquired abstract/metaphorical (grammaticalized) meanings, serving the same kind of functions as inflectional cases (Huddleston – Pullum 2002: 601). That is, the process of grammaticalization involves a shift in status of the preposition from a less grammatical to a more grammatical function (Kuryłowicz 1965: 69), including their metaphorical usage.

With concrete meaning, one may say that *at* is already there, it is precise in time and place: *at 5 o'clock, at the* door, *(I'll meet you) at the crossroads, (I am) at the sea, he hasn't arrived at the station yet*, etc., and can be analyzed as being deictic when it substitutes for *here* or *there*, e.g. *at home, at work.* In example (1), *at window* can be interpreted as *at the window*:

(1) I will go before sir.
Mistress, look out at window for all this, –
There will come a Christian by
Will be worth a Jewes eye.
(Shakespeare, *The Merchant of Venice,* II, v, 40)

The presence of the determiner (the definite article *the*) in the gloss signals that the preposition is referential for it has an identifying function. It is interesting to note, moreover, that the *at* complementation can have a resultative meaning, thus expressing the effect of the motion (cf. *for all this – There will come a Christian by*). Similarly, the sentence *we didn't see you at the party*, in which *at* heads a phrase of position, is the result of *you didn't come to the party* which indicates a destination.

With abstract (grammaticalized) meaning, *at* indicates the relation between emotion and stimulus: *I am surprised at his behavior, I am mad at him, I was annoyed at myself (for not exposing the problem), he is good/bad at mathematics,* etc. The following example exhibits a metaphorical meaning:

(2) No attempt was made at improvement, for they neither knew how to set about it nor could have gone to the smallest expense if they had, … (*Inverness Courier*, Aug. 1845)

In this sentence *improvement* is not a place or a position, therefore the preposition is used metaphorically to indicate a result which, according to the context, was not seen (resultative meaning).

Furthermore, *at* is also used to express an activity: *look at something, laugh at somebody, throw something at somebody, etc.* The verbs commonly used with this pattern (*aim, frown, gawk, glare, grab, grin, growl, hint, howl, laugh, leer, look, rush, shoot, shout, slap, smile, snatch, stare, swear, swing, wink, yell, …*) suggest 1) that *at* reinforces the idea that the reference point is a target: *aim at the bull's eye, the teacher smiled at the new student*, and 2) that it is linked with "look" and semantically related with descriptive verbs (*stare, glare, gaze, gape, wink, …* Lindstromberg, 1998: 168). Some of these verbs (e.g. *grab, shoot, snatch, …*), however, can be used either transitively or intransitively, with a prepositional phrase, which entails a semantic difference. Compare:

(3) a. The youth snatched the old woman's purse (and ran away with it).

b. The youth snatched at the old woman's purse (but he could not take it).

c. The hunter shot the deer in the head (the deer is dead).

d. The hunter shot at the deer's head (the deer managed to escape).

In the transitive case (3a-c) the action is effective and achieved, while the intransitive construction (3b-d) alludes to an unsuccessful attempt.

To, on the other hand, indicates orientation towards a goal: it is not yet there, it is to be achieved (it may include an indication of a distance, concrete or metaphorical). This can be seen in expressions like: *go to* (locative, concrete meaning), *object to something, similar to, look to/look forward to, see to* (abstract meaning)*: I'll see to it that everything's all right, Will you see to the outdoor chores? Give something to somebody* (dative), for example:

(4) Later Apple has quietly handed out refunds to unhappy owners. (*Newsweek*, Aug. 2010)

Note that some of the *to* complements can be construed as causes in relation to the resultative meaning of *at* phrases: compare *He went to the party* and *He was at the party*: *He was at the party because he went to the party*. The *to* complementation plays a causative role and thus represents an aim; it is also deictic and can be replaced by *there*: *because he went there*.

Since *at* is associated with the idea that the landmark is a target and *to* an indication of an aim to meet, it follows that there must be a difference in terms of orientation and reception.

4. Aim/target: What is the difference?

"An aim or purpose that you want to achieve is like a place that you want to get to or a target that you want to hit." This quotation, taken from the *Macmillan English Dictionary* (2nd edition 2007: 34), illustrates aptly the semantic implications of *aim* and *target.*

Aim is ambiguous (cf. *a place that you want to get to or a target*) and can appear with either *at* or *to*: *aim at/to, aim a gun at somebody, a reform aimed at reducing our expenses, aim to do something,…* A target, on the contrary, receives missiles: *you manage to get your shots on target, nuclear missiles will no longer be targeted on cities.* The preposition *on* confirms the idea of touching a surface or an object (cf. *a target that you want to hit*). There seems to be a correlation between *at* and *on* in the following:

(5) I have climbed Ladhar Bheim. […] The view was glorious. And I threw a banana skin at it. I have stood on the magnificent Aonach Eagach ridge and gazed down on Loch Achtriochtan. And I threw a banana skin at that; too. In fact, there are few mountains in Scotland haven't thrown a banana skin on. (*The Guardian*, Sept. 2009)

In example (5), it is possible to replace *on* by *at* in such phrases as …*and gazed down on (at) Loch Achtriochtan* and …*a few mountains I haven't thrown a banana skin on (at)*. There may be a difference in precision, *at* being more direct than *on* (cf. *about, around*) in that context, owing to the fact that *at* makes the target a one-dimensional location while with *on* the mountain becomes a two-dimensional area (Quirk – Greenbaum 1973: 147). In addition, *on* suggests that the area is seen from above (cf. *gazed down*) and that only part of it is referred to; *at*, on the contrary, implies considering the whole thing. It may be noted that *on*, besides expressing place (*on Loch Achtriochtan*) and time (cf. *on Tuesday*), can also indicate direction (cf. *on the left*: in the direction of the left), destination (*He fell on the ground*), and be found with noun phrases denoting ongoing actions or states (*on fire*, *be on drugs*, *on a diet*). Metaphorically it expresses reason (*on your advice*), a disadvantage as in *he died on me* (Huddleston – Pullum 2002: 661-2), abjurations (cf. *on my life, 'Upon my reputation and credit'*, Shakespeare, *All's Well that Ends Well*, IV, iii, 130),[3] or criticism as in (6):

(6) My family is always on me that I need to go back to college and get a real job. (Lady Gaga, *Newsweek*, June 6, 2011)

In the latter example, *on* can be analyzed as being close to *at* (cf. *my family is always nagging at me*) in that *me* is regarded as a target.

It is easy to see in what has been said in the previous lines that some of the additional meanings of *on* may be linked to, or clarify, the properties and interpretations carried by *at* (mainly position, destination, and the metaphorical uses of *on*). *On* also combines with *to* in such phrases as *be onto a good thing/something* or *be/get onto somebody* (e.g. *He knew the police were onto him for this crime*).

According to Boulonnais (2008), *to* expresses telicity, i.e. tending towards a goal envisaged as realized in a perfective sense, it involves an end-point (quantized). Does this imply that in the meantime *at* would be atelic in the sense that without it the *at*-complement would be incomplete? Of sentences (3a-d), for instance, (3a-c) appear to be telic, that is completely affected by the situation as presented by the speaker, while examples (3b-d) are only partially affected in the situation and may therefore be regarded as atelic. Yet, it appears that *at* phrases can be ambiguous given the context in which they occur and the semantic content of the predicate. The meaning of

3 *Upon* is a formal equivalent to *on*.

throw in example (5) shows that the *at* complementation serves an implicit purpose, so there must be something else that comes into play to distinguish what the two prepositions cover.

As a matter of fact, *to* and *at,* I argue, denote motion towards the same reference point, but seen from a different angle. Compare (7a-c):

(7) a. Everybody's talking at me, I can't hear a word they're saying. (*Macadam Cowboy)*

b. Are you talking to me?

c. He threw a stone at the dog.

d. He threw a bone to the dog.

e. Kate smiled to herself, and opened her eyes to see the anxious faces staring at her. (a fiction)

In (7a), *talk at me* means that I am treated simply as a target and I am commenting on the reception (cf. *I can't hear a word they're saying)*, while in (7b) *talk to me* suggests that the talker is communicating with me. Note that the interrogative form lays the emphasis on the origin of the communication. As Lyons (1977: 755) has rightly pointed out, "when we pose a question, we merely give expression to, or externalize, our doubt", and in so doing we establish a direct relationship with the addressee. This relationship is a cause-and-effect one since an interrogation naturally expects and may obtain an answer. In that, *at* and *to* can be legitimately opposed. The question, *Are you talking at me?* would sound unnatural to an addressee as the origin of the communication cannot be the actual target. In the prepositional phrase *at the dog* (7c), *dog* is the target, the subject (*he*) wants the dog (i.e. the receiver) to go away, he expects a result; *to the dog* in (7d) indicates the direction and implies the sender's intention. He wants to play with the dog. It involves his point of view.

A schematic representation of the implications of *at* and *to* could be thus:

(8) a. At: $x < y$

b. To: $x > y$

(8a) suggests that the target is seen from the receiver's viewpoint (cf. *arrive at, at a distance: from a place near here* as opposed to *in the distance, at the center* different from *in the center,*[4] expressions like *at last, at all, at least* which relate to the speaker or the receiving end, *be at it* meaning *doing something here and now*: W*hy don't you polish my shoes too, while you're at it?*), and in that sense the *at* phrase is telic since the event is presented as having an end-point. In (8b) the goal to achieve is seen from the sender's position. In example (7e), Kate produces the smile (*to*) and she receives the staring (*at*).

Now the idea of staring at somebody takes us back to the beginning of the story, back to the association of *at* with hostility, although *stare* does not necessarily imply unfriendly behavior. Furthermore, hostility may also be associated with *to* in the sense of "opposition to something": *There has been some hostility to the new conservative government*.

5. *At* and hostile intentions

I will end this discussion with a few more words about *at* associated with hostile intentions as shown in expressions like *laugh at, kick at, stab at, pull at, pluck at, shout at*. These are verbs which imply some kind of reception and their semantic dimension must not be ignored. Compare:

(9) a. He threw a ball at my face.

b. He threw a ball to his partners.

Sentence (9a) can be interpreted as hostile: I receive the ball smack dab in the face, while (9b) does not convey any hostility (cf. *partners*). The subject (*he*) sends the ball. The noun phrase, *his partners*, is not a target.

However, this idea of hostility is contradicted by expressions like *smile at*, often labelled as an exception. Although exceptions are said to confirm a rule, it would be irrelevant to apply such a statement to the use of prepositions as the latter are semantically vacuous: they are function words.

In fact, the hostile intentions depend on the semantic content of verbs (*kick, stab, shout*, *throw,* etc.) which generally imply reception, and/or on the contexts (*throw a stone* vs. *throw a bone, partners, …*) in which those

[4] *In* refers to either an area (in the center) or a volume (in the distance), and as such can be respectively two or three-dimensional. The notion of size may be taken into account: *at* is a spot *in* a larger space.

verbs occur. Some verbs, like *shout*, are ambiguous and may combine with *at* and *to*. With *to*, one tends towards a goal in order to establish a relation (communication):

(10) a. He shouted at me.

b. He shouted to me not to cross the line.

In (10a) I receive the shouting, and only the context will determine whether the subject's intentions are hostile: he may be angry with me, afraid or in pain: *the patient shouted at the nurse in pain*, but he may also shout because the receiver is hard of hearing, or will not hear. In (10b) the subject is telling me something in a loud voice.

When the verb is semantically unmarked, that is only denoting direction, the interpretation of *at* and *to* will depend on the different contexts. Examples (11a-b) include the use of the verb *direct* with both prepositions. But while (11a) clearly shows that the *at* complementation can be regarded as a target (cf. the context: *but I would direct my ire…*), sentence (11b) deals with the sending of a signal, thus implying an aim to achieve:

(11) a. What a vicious attack on Rebekah Brooks! Maybe she deserves it – I'm not a Brit, so she's new to me – but I would direct my ire **at Rupert Murdoch himself**. He's the one who hoisted gutter journalism to its current prominence. (*Newsweek,* the Mail, Aug. 15, 2011)

b. The news [of the resignation of the entire Turkish high command] shook Turks, but not Turkey's friends abroad **to whom** the military SOS signal was primarily directed. (*Newsweek*, Aug. 15, 2011)

In spite of the fact that in (11a) the speaker's meaning is somewhat hostile, his hostility is conveyed neither by the verb, nor by the use of the preposition *at*, but it can be inferred from the context: according to the speaker the attack is on the wrong target.[5] In (11b) the *Turkish high command* is, as it were, communicating with *Turkey's friends abroad* which can be construed as an end-point (telicity). The use of a passive construction, moreover, gives the status of topic to the direction (*to whom = to Turkey's friends abroad*) and the achievement of the objective.

5 Again, note the closeness of *at* and *on* in this example (cf. section 3).

6. Conclusion

As we have seen *at* and *to* are locative markers, the meaning of which is inferred from the context and the way the act is viewed. Their usage does not depend on collocation (*look at/listen to*)[6] or syntactic/semantic constraints (descriptive verbs, for example).

The receiving-end hypothesis then may explain why *at* is only followed by a noun or an -ING form (nominalization) which are related to stability, while *to* remains ambiguous, unstable,[7] relevant to the idea of a hiatus between the trajectory and the landmark, something that has not been reached (Cotte 1982a and b), and which can precede a verb, thus acquiring a metalinguistic value.

REFERENCES

Bottineau, D.
2002 "Les cognèmes de l'anglais : principes théoriques". In: R. Lowe (ed.) *Le système des parties du discours, sémantique et syntaxe*. Actes du IXe colloque de l'Association internationale de psychomécanique du langage. Québec: Université Laval, 423-437.

Boulonnais, D.
2008 "Prepositions TO and FOR: grammaticalisation et subjectification", *Anglophonia* 24, 45-97.

Bourquin, G.
1990 "De l'infini au défini : un aspect de l'organisation du sens grammatical", *L'organisation du sens CIEREC* (Travaux) 68, 105-123.

Cotte, P.
1982a "Autour de TO", *CIEREC* (Travaux) 35, 57-80.
1982b "TO opérateur de dévirtualisation en anglais", *Modèles linguistiques* 4 (2), 135-149.

Croft, W.
2010 "The origin of grammaticalisation in the verbalisation of experience", *Linguistics* 48, 1-48.

Freeborn, D.
1998 *From Old English to Standard English. A Course Book in Language Variation across Time* (2nd edn.). London: Macmillan.

6 *Listen at* and *look to* are possible occurrences.

7 From a cognemic point of view, the vocalic formant /U/ signals uncertainty, a potential state of the interval between two entities whereas /A/ denotes separation and distance (Bottineau 2002).

Hoad, T. F. (ed.)
1980 *The Concise Oxford Dictionary of Word Origins*. London: Guild Publishing.
Huddleston, R. – Geoffrey K. Pullum
2002 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Jespersen, Otto
1933 *Essentials of English Grammar*. [Repr. 2002] London: Routledge.
Kuryłowicz, Jerzy
1965 "The evolution of grammatical categories", *Diogenes* 51, 55-71.
Lapaire, J. R. – W. Rotge
1991 *Linguistique et grammaire de l'anglais*. Toulouse: Presses Universitaires du Mirail.
Lindstromberg, S.
1998 *English Prepositions Explained*. Amsterdam & Philadelphia: Benjamins.
Lyons, John
1977 *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
Macmillan English Dictionay for Advanced Learners (2nd edn.).
2007 London: Macmillan.
Meillet, Antoine
1912 "L'évolution des formes grammaticales". In: A. Meillet *Linguistique historique et linguistique générale*. Paris: Champion, 131-148.
Mitchell, Bruce
1985 *Old English Syntax*. Vol. 1. Oxford : Clarendon Press.
Quirk, Randolph – C. L. Wrenn
1957 *An Old English Grammar* (2nd edn.). London: Methuen.
Quirk, Randolph – S. Greenbaum
1973 *A University Grammar of English*. Harlow: Longman.
Random House Webster's Unabridged Dictionary.
2001 New York: Random House.
Rapatel, Ph.
2010 "Le rôle des prépositions: le cas de FOR", *Anglophonia* 28, 11-32.

Devil *aka* Satan: An *enemy* or *fiend*? On the rivalry between the familiar and the foreign in early English

Joanna Esquibel and Anna Wojtyś
University of Warsaw and University of Warsaw

ABSTRACT

The present paper discusses the distribution of the two most common Mediaeval English euphemisms of Satan, i.e. *fiend* and *enemy*, in religious prose. We focus on the rivalry between the foreign word and the native word, comparing the contexts in which the two words tended to occur, and attempting to determine the semantic status of the French word *enemy* in relation to the sense originally denoted by *fiend*. The data come from the Middle English period, when French loanwords began to compete semantically with native words.

1. Preliminary remarks

The history of English shows taboo-induced replacement (cf. Hock – Joseph 1996: 232) of the words referring to positive and negative supernatural powers. As stated by Hughes (2000: 44) "the motive is to describe the situation as better than it is, or to avoid the taboo area, thereby pacifying some dreaded force by managing not to offend it…". Thus, the use of euphemisms may be a result of fear and/or respect, especially in superstitious times (cf. Katamba 2005: 191), leading, for instance, to the substitution of religious proper names such as Jesus or Satan by their euphemistic synonyms (cf. McMahon 1994: 181), e.g. *haelend* 'healer' or *witherwin* 'adversary', respectively.

Thus, not without reason, in mediaeval England, the evil powers, i.e. Satan and his followers, the devils, were referred to by a whole variety of euphemistic expressions. The present paper focuses on two euphemisms which gained prominence in Middle English, namely the Germanic *fiend*

and the newly-borrowed Romance *enemy,* presenting the circumstances of the first attestations of the borrowing and its temporal and dialectal spread in religious writings of the period. The distribution of *enemy* is further compared with that of the native *fiend* to verify whether the two items were employed in the same contexts. The study is expected to reveal the place of the two words in the semantic domain of SATAN and, thus, suggest the plausible reasons for the introduction of the foreign element, be it the result of need or, rather, prestige (cf. Campbell 2004: 64).

2. Euphemisms for Satan

When it comes to Satan in English mediaeval texts, the creature is rarely addressed directly by its proper name. The *Historical Thesaurus of English* (henceforth referred to as HTE) provides a number of semantically varying synonyms which were used instead, cf.:

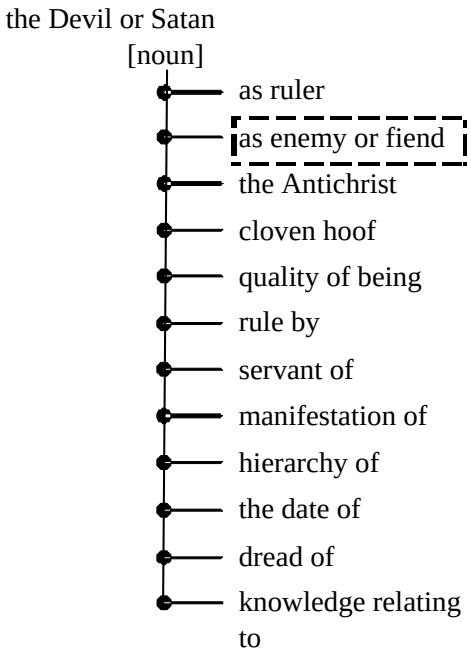


Figure 1. The categories of the Devil or Satan from HTE

To narrow the scope of the data, however, the present study covers only the semantic category of Devil or Satan in the meaning of 'enemy or fiend'.

The Old English terms provided by the HTE within that category include *witherwin* (c897) and *fiend* (a1000). Further on, in Middle English, the next three euphemisms attested are *our foe* (?c1225), *fed* (a1300) and the first two foreign words, i.e. *adversary* (1340/ 1667[1]) and *enemy* (1382). Later synonyms include forms intensified by the prefix *arch-*, i.e. *arch-foe*, *arch-traitor*, and *arch-enemy*.

The HTE list of euphemisms is by no means complete. Other historical dictionaries, such as Bosworth–Toller's *An Anglo-Saxon Dictionary* (henceforth BT) and the *Middle English Dictionary* (henceforth MED), provide further terms, such as OE *wiþerbreca*, *wiþerhycgende*, *witherweard*, *withersaca*, and ME *unwine*, all of which had the meaning of 'enemy' (cf. OED, MED) and were used with reference to Satan. For the expanded list of items, see Table 1.

Table 1. Synonyms of the Devil or Satan (based on HTE, OED and MED)

Period	Item	First attestation
OE	*wiþerbreca*	OE
	wiþerhycgende	OE
	wiþerweard	c888
	wiþerwin	c897
	feond	a1000
	wiþersaca	a1150
ME	*oure fo*	?c1225
	unwine	a1225
	fed(e)	a1300
	adversārie	1340
	enemī	1382
ModE	*arch-foe*	1667
	arch-traitor	1751
	arch-enemy	1850

Interestingly, most of the Old English words are complexes formed with the prefix *wither-* 'against, in opposition' (OED) attached to the nouns, thus adding the negative meaning of hostility.

1 Although the HTE dates the first attestation of *adversary* in the religious meaning to 1667, both the *Oxford English Dictionary* (OED) and the *Middle English Dictionary* (MED) quote an earlier example of its use in *The Ayenbite of Inwyt*, a holograph dated to 1340. Thus, that date is taken into consideration in the present study.

For the sake of space and time, however, the present analysis is further narrowed down to the two most prototypical euphemisms in Middle English religious jargon referring to Satan, namely native *fiend* and foreign *enemy*, which, as Fig.1 shows, are quoted in the HTE as two synonyms of categorically equal semantic content.

2.1 *Fiend*

The noun *fiend* derives its meaning from the present participle form of the verb *fēogan* 'to hate' and as such it refers to an enemy or hater (Skeat 1968). Originally spelled *feond*, in Middle English the word had numerous orthographic variants, including (but not limited to) *feond*, *fende*, *finde*, *feynd*, *fynde*, *feende*, and *fiende*.

Fiend is attested in English for the first time in the 10th century *Rushworth Gospels*, in the phrase *Hate þine fiond* (Matt. v. 43, c975) and shortly afterward it starts to be used in reference to 'the arch-enemy of man-kind, the devil' (OED) or 'Satan' (MED), with earliest attestations before or around the year 1000, cf.:

(1) a1000 Ðu **fiond** geflæmdest. (*Hymns* (Gr.) viii. 25)
c1000 Hit eac deah wiþ **feondes** costungum yflum. (*Sax. Leechd*. II. 294)

In addition to the purely euphemistic sense of 'Satan himself', use of the word here also denotes 'an evil spirit generally; a demon, devil, or diabolical being' (OED), cf.:

(2) OE No þær þa **feondas** gefeon þorfton. (*Guthlac* A 421)
c1175 Ah a þer is waning and graming … and **feonda** bitinga. (*Lamb. Hom.* 33)

This use often refers to the Devil as one of the evils tormenting man, cf. þreo cunne uan: þe ueont & teos wake worlt ant hare licomes lustes (St.Marg. (Bod 34)).
In time, the term's meaning broadened to that of 'a person of superhuman wickedness' (OED), cf.:

(3) c1220 For wo so…ðenkeð iuel on his mod fox he is and **fend** iwis. (*Bestiary* 450) c1300 He with his hend Ne drop him nouth, that sor **fend**. (*Havelok* (Laud) (1868) 2229)

Typical collocations quoted in historical dictionaries also indicate religious connotations with either Satan or his followers and other evil spirits. The most frequent are *the fiend of hell*, attested as early as c1225, and *foul fiend*, referring to both Satan and devils. Others usually indicate either the position or importance of Satan or the Devil, including *the heigh fiend* 'the Arch Enemy' or *the old fiend* 'the Ancient Foe', or kinship, †*fiend's limb*, *kin*, *child*, etc. Not without reason, when describing the Devil, mediaeval scribes would often refer to its devilish qualities, in collocations such as *envious fiend*, *fals fiend*, *fiend unfre*, *wikked fiend*, or *wrenchful fiend*.

2.2 *Enemy*

The word *enemy* originally comes from Latin *inimīcus*, 'unfriendly', formed of the negative prefix *un-* attached to the adjective *amīcus.* It entered English via French, as *enemi* or *anemi*, in the early 14th century. Characteristically, in the Middle English period, *enemy* occurs in various spellings, the most frequent of which are *enemi(e*, *enemy(e*, *enmi(e*, and *enmy(e*. Additional, quite unconventional orthographic variants are also found, the most extreme examples being *elmy* or *elmee*, which are recorded in the letters of the Paston family.

The first attestation of English *enemy* mostly likely dates to the turn of the 14th century (cf. the appearance of the noun in MS Cambridge University Library Gg. IV.27 (2) of c1300 (?1225), which contains *King Horn* and a portion of the *Cursor Mundi*. The MED provides the plural form *enemis* in a quotation from that MS version of the former text.

The general meaning of *enemy*, i.e. 'one that cherishes hatred, that wishes or seeks to do ill to another' (OED), allowed the word to develop more specific meanings such as that of 'a member of a hostile army' and 'a destructive quality or force' (OED, MED). Within a hundred years of its assimilation into the English lexicon, *enemy* also started to be used with reference to Satan, the first attestation coming from the early version of the Wycliffite Bible, dated to 1382 (OED), cf.:

(4) I haue ȝouun to ȝou power of defoulinge, other tredinge on… al the vertu of **the enemy**. (*Bible Wycliffite, E.V.,* Luke x. 19)

It is interesting to note that the MED quotes the passage from Chaucer's *The Tale of Melibee* as the first use of enemy in that meaning, but the dating provided (c1390) is that of the original not the manuscript. Moreover, within

the entry of *enemy* defined as "Of an evil spirit, esp. the Devil", the dictionary does not provide any quote from the Wycliffite Bible.

According to these dictionaries, in the religious meaning, the word *enemy* most often collocates with possessive pronouns and adjectives such as *ghostly*, *great*, and *old.* Other frequent collocations are seen in the phrases *the enemy of hell*, *the enemy of mankind*, and *the enemy of souls*. The noun is typically preceded by the definite article *the* suggesting reference to Satan himself.

3. Data

The data for the present study come from *The Innsbruck Corpus of Middle English Prose*, which contains complete Middle English prosaic texts of various genres. The use of these texts proved essential, since, unlike poetry, where the selection of words is often determined by metre and rhyme, prosaic data provide a more reliable source of information. As the dating and the dialectal distribution are of importance for the research, the study includes only those texts which come from manuscripts of fairly reliably specified dates and provenances. Thus, the list of texts examined includes 59 sources from all five Middle English dialects, dated to various centuries of Middle English, ranging from a1200 to 1500. An additional advantage of that text selection is that it represents various genres (including that of chronicles, etc.), which allows for a more thorough analysis since religious meanings and contexts are also found in secular texts.

All the texts have been examined for forms of the two words central to the study, i.e. *fiend* and *enemy*. From all the instances of their usage, those with religious meanings were singled out in order to establish their frequency in various periods of Middle English. Furthermore, the contexts in which these two nouns refer to Satan, the Devil, or a devilish creature were compared in order to discover potential differences in their applications.

The study disregards those texts where neither of the nouns appears (12 texts)[2], and it focuses out of necessity on those where at least one instance of either *fiend* or *enemy* is evident. The textual sources examined are divided into three categories according to the presence or absence of each noun. There are those in which:

---

2 i.e. *Twelfth-Cent. Homilies* (Bod 343), *History of the Holy Rood-tree*, *Old English Homilies*, *Vices and Virtues*, *Kentish Sermons*, *Two Fifteenth-Century Cookery Books*, *Agnus Castus. A Middle English Herbal*, the works of John Metham: *Christmas Day* [1], *Christmas Day* [2], the *Days of the Moon*, *Palmistry*, and *Physiognomy*.

(1) *enemy* is absent and *fiend* is present;
(2) *enemy* is present and *fiend* is absent;
(3) *enemy* is present and *fiend* is present.

For the purposes of the present analysis, the last group, including both items, is of greatest significance and, as such, is discussed in greater detail below.

3.1 The absence of *enemy* and the presence of *fiend*

The first category takes in the texts that contain instances of *fiend* but not *enemy*. As should be expected, the majority of those sources are dated to the period before *enemy* had, to our knowledge, entered the language, i.e. the first centuries of Middle English times, cf.:

Table 2. Texts including *fiend* but not *enemy*

Date	Dialect	Text
[1150-1250]	WMdl	*Seinte Marherete* (Roy)
[1150-1250]	WMdl	*St. Julian* (Bod)
[1150-1250]	WMdl	*St. Julian* (Roy)
[1150-1250]	Kentish	*Twelfth-Cent. Homilies* (Vesp)
1150-1250	WMdl	*Hali Meidenhad* (Bod)
1150-1250	WMdl	*Hali Meidenhad* (Tit)
1150-1250	WMdl	*Hali Meidhad* (crit)
1150-1250	WMdl	*Sawles Warde*
12/13c.	WMdl	*Wohunge of Ure Lauerd*
?a1200	Southern	*Ancrene Riwle* (Ner)
?a1200	WMdl	*Ancrene Riwle* (Tit)
?a1200	WMdl	*Ancrene Wisse* (Corp-C)
?c1200	WMdl	*St. Katherine* (Roy)
a1225	EMdl	*Old English Homilies*
c.1200-1250	WMdl	*Ancrene Riwle* (Gon-Ca)
c1230	WMdl	*Seinte Marherete* (Bod)
1340	Kentish	Dan Michel, *Ayenbite of Inwyt, or Remorse of Conscience*

This class also includes texts that originated in Old English and were copied only later, such as the so-called AB language works. The chronologically most recent text to exhibit *fiend* exclusively is Dan Michel's *Ayenbite of Inwyt,* a Kentish translation of *Somme le Roi*, dated to 1340. Given the fact that Dan Michel is known for having translated the French source *literatim*, thus including direct calques from French (Janecka 2008: 151), it is interesting that the French borrowing *enemy* is not attested in his work, as it is claimed to have entered English at least half a century earlier (see section 2.2). The absence of the word might be attributed to what has been called "the conservative nature of his language" (Gradon in Morris 1965: 12), which represented not the mid-14th century but rather the late 13th century, (Laing 1993: 67), the time when *enemy* should not have been widely recognized yet.

In the texts examined, the word *fiend* is employed in more than a single meaning. As the selected data illustrate, it was an antonym to *friend* (5a), but it also denoted 'an opponent' in general (5b) or 'an opponent in a fight' in particular (5c), cf.:

(5) a. þine **frend** sariliche wið reming and sorhe. þine **fend** hokerliche to schome and wundren up o þe. A nu haue þai broht him þider. (*Wooing Lord,* p. 283)
Mi feader & Mi moder for þi þ ich nule þe forsaken: habbe forsake me. & al mi nestfalde cun. þ schulde beo me best **freond** beoð me meast **feondes**. (*St. Julian* (Bod), l.277-280)

b. seo swicola Dalila þone strange Sanson, hire agene were, mid olæcunge bepæhte, & bescorene fexe his **feonden** belæwde. (*Twelfth-Cent. Homilies*, (Vsp)., p.57, l.12-14)
[…] to uoryeue þe on to þe oþre and louye oure **uyendes** […] (*Ayenbite of Inwyt,* p.114)

c. hwen me asaleð burhes oðer castel; ţeo wið innen healdeð scaldinde weater ut. & werieð swa ţe walles. ant 3e don alswa as ofte as ţe **feond** asaileð ower castel & te sawle burh […] (*Ancrene Wisse*, p.125, l.1-5)

Still, the sense in which *fiend* is used most frequently is one involving evil power, or, more specifically, one synonymous with Satan. As such, the noun is found especially in religious texts, such as sermons, homilies, or texts directed at members of religious orders. The incidence of that usage might be indicated by the high frequency with which the word occurs in that sense in the text *Ancrene Wisse* (Corp-C 402); 66 instances of such usage appear there.

3.2 The presence of *enemy* and the absence of *fiend*

The database also includes texts that utilize *enemy* but never *fiend*. It must be noted that all the texts here are dated to the 15th century, when the French word is assumed already to be well-rooted in the language. As Table 3 shows, it was especially common in the eastern areas, cf.:

Table 3. Texts including *enemy* but not *fiend*

Date	Dialect	Text	*enemy*	
			Total	Religious
[c1400]	WMdl	*Brut, or The Chronicles of England*	61	---
c1425	EMdl	*Fistula in ano*	2	---
1420-1500	EMdl	*Paston Letters*	93	---
c1452	EMdl	Capgrave's *Lives of St. Augustine*	12	---
a1464	EMdl	Capgrave's *Chronicles, Abbreviation* of	38	---
c1450	EMdl	*Secreta Secretorum*	10	---
1472-1488	EMdl	*Cely Letters*	1	---
a1475	EMdl	*Spheres and Planets*, in *The Book of Quintessence*	1	---
a1500	EMdl	*Secreta Secretorum*	12	---
c1400	Southern	*Three Middle English Sermons* (Worcester Chapter Manuscript F. 10)	28	5
a1450	EMdl	Trevisa, *Methodius, The Bygynnyng of the World*	2	2

Interestingly, in the majority of those texts the loanword is used in the secular meaning only, referring to notions other than Satan/the Devil or devils. This, presumably, follows from their genre since they are mainly non-religious documents such as medical or astrological treatises, historical chronicles, or private letters. The word *enemy* is employed here in the meaning of someone opposing something or, more specifically, an armed opponent in a fight (6a). Also, the meaning of the word is sometimes metaphorical, expressing a destructive or hostile force (6b-c), cf., e.g.:

(6) a. […] þei spedde faste toward her **enemyes** for to 3eue hem bataylle […] (*Brut,* p.12, l.33-34)

b. ffor after ypocras cold þing3 in acte bene **enemys** to bone3, to synowe3, to teþe, to brayne, to þe lure, to þe bladdre, and to þe nerw3 of þe rigebone. (*Fistula in ano,* p. 70, l. 22-25)
c. Mars is an **enemy** to alle thyngis to be gendrid; wherfor he is clepid god of batel […] (*Spheres and Planets*, p. 26)

In the two texts which employ *enemy* exclusively, i.e. a treatise by Trevisa entitled *Methodius*, written in the East Midland dialect, and *Three Middle English Sermons* in a Southern manuscript (Worc F. 10), the word is used in the religious sense. In *Methodius, enemy* occurs twice with reference to the Antichrist (7a). In the *Sermons*, only five out of 28 instances of the word reflect religious usage. Curiously, in most passages where the word denotes the Devil, the referent is further specified, cf. (7b-e):

(7) a. And onone he sal sla þis beste, Antecrist **enmy** & disceyfer, with þe swerd of his mowthe […] (*Methodius*, p.111, l.22-24)
b. […] I vndirstond no-thyng ell at this tyem bot hour gastlyche **enmy, þe deuel of hell**, þis kursyd i3e, þis wyckyd Pharoo […] (*Three Middle English Sermons*, p.23, l.31-33)
c. […] & lift vr sowle fro þe stynkynge dingel o lustes o þis world, þat **vr enmy, þe deuel of helle** […] (*Three Middle English Sermons*, p.58, l.255-256)
d. […] 3if we þus do; **vr enmy, þe deuel**, schal ner take a-way vr offryng' from vs. (*Three Middle English Sermons*, p.65, l.477-479)
e. […] whan a streþte His bodi o þe cros & þrew doun mannis **enmy, þe deuel of helle** […] (*Three Middle English Sermons*, p.49, l.884-885)

Because the majority of uses express a secular meaning of *enemy* as the opponent, it seems that in the constructions quoted above additional phrases such as *the devil of hell* function as an explicitation of the religious sense of the term. This, in turn, suggests that the religious meaning of *enemy* might then have been considered peripheral.

3.3 The presence of both *enemy* and *fiend*

For purposes of the present study, of greatest importance are the texts which contain both words, *enemy* and *fiend*, since they might display differences in the employment of the two items. Some of the texts include the two euphemisms in both secular and religious meaning, cf.:

Table 4. Texts including both *enemy* and *fiend* in various meanings

Date	Dialect	Text	*enemy*		*fiend*	
			Total	Religious	Total	Religious
a1425	Emdl	*Adam and Eve*	2	2	7	7
1434	Emdl	Misyn, *The Mending of Life*	3	1	11	11
1435	Emdl	Misyn, *The Fire of Love*	20	5	27	27
?a1450	North	*Alphabet of Tales*	27	1	150	150
a1450	Emdl	*Pater Noster of Richard Ermyte*	5	3	14	14
?c1450	Southern	*The Book of The Knight of La Tour-Landry*	15	5	17	17
c1450	Emdl	Julian of N.'s *Revelations* (Shorter Version)	5	5	16	16
a1500	WMdl	*De Imitatione Christi*	26	8	4	4
a1500	WMdl	*Speculum Sacerdotale*	28	2	2	2
a1500	Kent	*Merlin*	105	5	29	29

In only two of the texts listed in Table 4, Julian's *Revelations* and *Adam and Eve,* do all instances of both words, often used in proximity, denote an evil power, cf.:

(8) a. And whanne Adam say hir, he cryede wepynge: "O Eue, where is þe werk of þi penaunce; how is it þat **oure enemy** haþ þus bigylid þee, […] "Whanne Eue herde þis, sche knew hir-silf bigylid þoru þe **feend**, and fel grouelynge to þe erþe […] (*Adam and Eve*, p.83, l. 16-21)

b. […] 3e, vnto alle creatures lyevande that schulde be saffe agaynes alle the **feendys** of helle & agaynes alle gostelye **enmyes**. (Julian's *Revelations*, p. 43, l.24-26)

c. For I trowe sothlye, ware I saffe fra synne, I ware fulle saife fra alle the **fendes** of helle & **enmyse** of my saule. (Julian's *Revelations*, p. 75, l.19-20)

While in item (8a) the terms seem to be nearly synonymous, items (8bc) allow for the determination that these nouns have a semantic scope broader than one encompassing spiritual opponents which are only *fiends of hell,* a phrase that refers specifically to devils. Also, in both sources, *enemy* is less frequent

than *fiend,* the ratio being 2 to 7 in *Adam and Eve* and 5 to 16 in *Revelations*, respectively.

In the remaining texts listed in Table 4, only the Germanic word is used in religious contexts. In contrast, *enemy* typically refers to secular opponents and only occasionally denotes Satan or devils, e.g.:

(9) Cesarius tellis of a knyght þat on a tyme was taken with his **enmys** & slayn. […] And when he dyed, a man þat was vexid with a **fend** was delyverd. (*Alphabet of Tales*, p. 331-332)

As item (9) shows, the word *enemy* denotes the knight's opponents who kill him in a fight. But when referring to the devil that possessed the man, the author employs the word *fiend*, not *enemy.* A similar distribution may be observed in other texts, which suggests that *fiend* rather than *enemy* tends to be associated with a religious meaning. Still, the two items are occasionally treated as being nearly synonymous, cf.:

(10) a. Happy is þe ryche þat has slike possessyon; & þis to haue þe warldis vanyte þou forsake: & he þe **enmy** sal ouercome & þe to his kyngdom brynge. Þe **feynd** sall be ouercomen þat þe noys, þe flesch made sogett þat þe greuys (Misyn, *The Fire of Love*, p.63, l.5-7)
b. But true mariage is ordeined be God […] and therfor the **fende** of hell hathe no pouer in that holy sacrement, […] and, as a smithe that is euer blowinge in the fire, and right so seruithe the **ennemy** of hell that besiethe hym euer forto kendill and lyght the fere flame of dedly synne witheinne the hertis of man and woman bi fals delite […] (*The Book of The Knight of La Tour-Landry*, p. 164, l. 28-32)
c. Witirly noon but þe envious **enemy**, þe **feende** of helle, þat euer ylike procuriþ wiþ his wrenchis […] (*Pater Noster*, p.6, l.19-20)

Interestingly, in the above passages both the native and the borrowed term happen to be used in the same collocations, cf. *The enemy shall overcome* vs. *The fiend shall be overcome* (10a), and *the fiend of hell* vs. *the enemy of hell* (10b).

Item (10b) further confirms that both words may occur with the same collocates, such as the adjectives *foul* or *envious*, or the postmodifier *of hell*. That phrase, however, most frequently modifies *fiend*, while the collocation with *enemy* is sporadic. Still, identical modification seems to suggest that the foreign item not only took on the semantic properties of the native *fiend* but also some of its collocates. On a side note, the fact that *enemy* is also often

preceded by determiners may indicate the need for more precision when referring to the Devil or his followers, hence *our enemy, his enemy,* etc.

The remaining sources that utilize both items, all of the East Midland dialect of the 15th century, show a clear semantic distribution of the two words in question, cf.:

Table 5. Texts including *enemy* in secular contexts and *fiend* in religious contexts

Date	Dialect	Text	*enemy*		*fiend*	
			Total	Religious	Total	Religious
a1400	EMdl	*Ancrene Riwle* (MS Pepys)	17	---	72	72
a1400	EMdl	*The Gospel of Nicodemus*	3	---	1	1
a1400	EMdl	*Pepysian Gospel Harmony*	5	---	56	56
a1450	EMdl	*Mandeville's Travels* (Bodley Version)	4	---	1	1
c1450	EMdl	*Speculum Christiani*	19	---	23	23
c1450	EMdl	Lavynham, *A Litil Tretys*	1	---	10	10
a1475	EMdl	*Book of Quintessence*	2	---	2	2

In all of the texts listed in Table 5, *fiend* is employed solely in its religious meaning, while the use of *enemy* is restricted to the secular use. This indicates a high degree of specialization of the two items in those works, and foreign *enemy* has acquired a secular meaning in certain instances. Compare the uses of the two words within the same texts in (11):

(11) a. As Iudas betrayede Criste to his **enmys**, so the mynister of the sacramente or the receyuer vnworthi, in as mych as in hym es, be-take[3] hym to deueles, whil he putte3 [hym] in a place that es vnder power of **fendes**. (*Speculum Christiani*, p. 178, l. 16-19)

b. þei putte awey also þe craft of **þe feendis** temptaciouns, and ymagynaciouns of dispeir. þei distroie, & make a man to for3ete almaner of yueles, and naturaly bryngiþ him a3en to resonable witt. and for as myche as saturne þe planete naturaly ys coold and drye, and is **enemye** to al kynde. (*Book of Quintessence*, p. 18, l.12-16)

c. Whan þe **deuel** assaileþ 3ou.casteþ out scoldyng water opon hym as men done att Castels opon her **enemyes**. For þere þat water comeþ. **þe fende** flei3eþ sikerlich. (*Ancrene Riwle* (Pepys), p.111, l.25-27)

In the items under (11), the two words appear in close proximity, which, we may suppose, the firmer difference in their meanings allows. In all three quotations, the word *enemy* denotes an opponent, either in a general sense (11ab) or in a military one (11c). In contrast, *fiend* always has a religious connotation denoting evil power.

It is interesting to note that the native word is used not only to refer to Satan himself but also to signify all kinds of devils. In such a context, it is not preceded by a definite article and may take various other modifiers, including the indefinite article (12a), a numeral (12b), or an adjective (12c). It may also assume a plural form (12d), cf.:

(12) a. Now was þere a man amonges hem þat hadde **a fende** wiþin hym. (*Pepysian Gospel Harmony*, p.19, l.27-28)
b. Hou þat Jesus enchasced **sex þousande & sex hundreþ and sexti & sex fendes**, and after passed hym ouer þe se. (*Pepysian Gospel Harmony*, p.20, r.19)
c. […] the synnes accusynge schal be on the ryght syde, **innumerable fendes** scha[l] be on the lefte syde […] (*Speculum Christiani*, p.54, l.3-4)
d. for **fendis** aperyn to hem opynly and afrayen hem and flyen into the eye with thondyr and fer and othere hidous tempestis (*Mandeville's Travels*, p.105, l.22-24)

The plural usage of the word may in part have been the result of a semantic extension which transferred the meaning of Satan to that of devils associated with him. Strangely enough, the word *devil* does not seem to have been treated as taboo; it was frequently employed in Middle English. The corpus of texts examined yields more than a thousand instances of its use. Thus, even though *devil* in its meaning of 'the supreme spirit of evil' (OED) may often have been replaced by euphemisms, the word in its other meanings, such as 'evil or unclean spirits' (OED), also started to lose ground to some of the same euphemistic expressions, such as *fiend.*

Elsewhere, the same word, *fiend,* may also be interpreted in a broader sense as a kind of embodiment of evil, i.e. an evil person or creature, cf.:

(13) a. And þo seide Jesus þat on of hem twelue was a **fende**. And þat he seide of Judas, þat hym bitraied. (*Pepysian Gospel Harmony*, p. 48, l.38-39)

b. And þis knyght held hur still, & þis womman pullid faste & wolde hafe bene away. So at þe laste sho pullid so faste at all hur hare braste of hur heade, & sho ran away & þis **fend** folowd after & tuke hur […] (*Alphabet of Tales*, p.310, l.20-23)
c. […] till that Gawein that to euery nede was nygh it a-parceyved, and saugh the grete harme that he dide of her peple, and seide to hym-self yef this **feende** lyve eny while we may moche lese. (*Merlin*, p.589)
d. […] the victorie that he hadde yeve the kynge, ffor neuer hadde | thei seyn so grete a **feende**; […] (*Merlin*, p.649)

As these quotations show, the word *fiend* was employed in the sense of 'a person of superhuman wickedness' (OED). Still, it could reasonably be argued that such persons were treated here as those who serve, or even personify, the devil. Hence, given the scarcity of data (4 cases), the present study classifies that sense as religious.

4. Conclusions

The analysis of the distribution of the two items under scrutiny yields the following conclusions:

(1) the semantic range of the native word *fiend* has taken in opponents of all kinds, yet the core meaning of that item seems to have been one of an opposing evil power;
(2) the French word *enemy* is attested in its initial English meaning of opponent in the prosaic texts of the database for the first time at the beginning of the 15th century;
(3) at a certain point, *enemy* broadened its semantic scope and began to be used in a religious sense as a euphemism for Satan/ the Devil, especially in East Midland and Southern (15th c.), and, occasionally, it also appears with collocates typical of the native *fiend*, cf. *enemies of hell;*
(4) however, *enemy* does not maintain its religious sense well, being attested rarely and sporadically (in 12 of the 59 texts examined) with a religious meaning – yet comparatively frequently with meanings such as 'opponent in life', 'opponent in battle', etc.;
(5) simultaneously, *fiend* loses non-religious meanings, and by the 15th century is used with reference to Satan and/ or his followers only;

(6) in time, the rivalry between the two words led to a quite clear semantic distinction, one which is still observed in Present Day English, i.e. the specialization of *fiend* to religious contexts, and the narrowing of *enemy* to the non-religious denotation of 'opponent'.

REFERENCES

Sources

Bosworth, Joseph – T. Northcote Toller (eds.)
1898 *An Anglo-Saxon Dictionary.* Oxford: Clarendon Press.
Innsbruck Computer Archive of Machine-Readable English Texts (ICAMET)
1999 Compiled by Manfred Markus (University of Innsbruck). CD ROM version.
Middle English Compendium: The Middle English Dictionary, A Hyper Bibliography of Middle English Prose and Verse, a Corpus of Middle English Prose and Verse Compiled by Frances McSparran (University of Michigan). Humanities Text Initiative. http://ets.umdl.umich.edu/m/mec/.
Oxford English Dictionary Online (OED online)
http://www.oed.com/.
Skeat, Walter William
1968 *An Etymological Dictionary of the English Language.* Oxford: Clarendon Press.

Special studies

Campbell, Lyle.
2004 *Historical Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.
Hock, Hans Heinrich – Brian D. Joseph
1996 *Language History, Language Change and Language Relationship. An Introduction to Historical and Comparative Linguistics.* Berlin: Mouton de Gruyter.
Hughes, Geoffrey
2000 *A History of English Words.* Oxford: Blackwell.
Katamba, Francis
2005 *English Words: Structure, History, Usage.* London: Routledge.
McMahon, April M.S.
1994 *Understanding Language Change.* Cambridge: Cambridge University Press.

Morris, Richard – Pamela Gradon (eds.)
1866/1965 *Dan Michel's Ayenbite of Inwyt*. (EETS 23). Oxford: Oxford University Press.
Janecka, Joanna
2008 *Periphrastic and Suffixal Adjectival Grading in Middle English (12th-14th c.): A Corpus Study*. (Warsaw Studies in English Historical Linguistics 2.) Warsaw: Institute of English Studies, University of Warsaw.
Laing, Margaret
1993 *Catalogue of Sources for a Linguistic Atlas of Early Medieval English.* Cambridge: D.S. Brewer.

Patient imaging in English medical case reports

Magdalena Murawska
Adam Mickiewicz University, Poznań

ABSTRACT

The present paper addresses the way patients are depicted in English medical academic texts aimed at health professionals. It examines the linguistic choices authors make and the effect they produce. The key issue is the construction of a patient persona, by analogy with other studies in which the authorial persona has been researched. In my project, I have attempted to analyse a corpus of selected case reports in search of references to patients in order to establish the roles they are assigned. Identifying these roles would help to determine patients' textual status, i.e. whether they function as the subjects or the objects of a medical study. The results of the analysis are discussed with reference to patient-centred approaches in medicine and to some facts concerning the history of medicine which have influenced the visibility of patients in medical literature.

1. Introduction

The language medical professionals use in order to document their academic activities has been widely researched in recent years within the framework of specialised discourse analysis. The bulk of the studies of written medical discourse constitute quantitative investigations into specific lexical and grammatical features (Gotti – Salager-Meyer 2006) and their respective functions as well as how authors organise and present ideas in specific text parts (Myers 1990). Another substantial body of research has centred on the broad theme of patient imaging. Under this umbrella term can be subsumed studies devoted to linguistic representations of patients, doctors and diseases. This research has addressed impersonality (Hyland 2001), authorial identity (the KIAP project) and metaphors (Sontag 1991) to mention but a few. Yet, the present author knows of no study of patient representation in medical

texts which considers the various contextual factors of their production. In this paper, I investigate patient imagining in a corpus of fifty medical case reports composed in Present-Day English in order to reveal the linguistic choices the text authors make as they write about patient diagnosis and treatment. I focus on the effect that particular selections produce and the possible factors influencing the language use patterns presented. I begin with an overview of the issues concerning written medical discourse, i.e. its characteristics and operational context. Next, the data and the methods applied in the study will be described. Finally, the results of the analysis will be discussed and conclusions will be drawn.

2. Theoretical background

The course by which to examine the relation between the form and the content of medical texts has been determined by the common practice to perceive their language as neutral, economical and depersonalized (Kenny – Beagan 2004: 1072), to a large extent due to the notorious use of the Passive Voice (cf. Albert 2004; Kenny – Beagan 2004). In general, these features are common to scientific discourse, yet, according to Bazerman (1988), scientific discourse is shaped by a given discipline (1988: 47). It follows that the ways authors give account of their scientific activities are influenced by modes of reasoning, methodologies and objectives of a given area of study. This hypothesis has been accepted by Taavitsainen and Pahta (2000), Atkinson (2001) and others who have examined relevant scientific papers. Following this line of reasoning, the features of medical texts might be conditioned by the nature of medicine as an area of study and practice. As regards the former, medical discourse may reflect the premises of the *biomedical model of medicine*, which has dominated since the 19th (c.) In short, this framework views illness as a direct consequence of the diseased body and patients as mere recipients of treatment (Wade – Halligan 2004: 1398). Regarding medical practice, according to Beagan (2000), in the course of medical training, students are taught to execute objectivity and personal withdrawal in their practice. They learn to report only facts and to limit personal input to a minimum (cf. Freidson 1970; Lock – Gordon 1988). The presented premises of the *biomedical model* have had to confront alternative models of medicine, for example *patient - centred medicine* and the *biopsychological model of medicine* (cf. Engel 1977). While the first of these "conceives of the patient as an experiencing individual rather than the object of some disease entity" (Mead – Bower 2000:

1089), the second advocates the incorporation of the patient's "whole self" (Wade–Halligan 2004: 1400) into the processes of diagnosis and treatment. If these postulates are considered in the context of medical texts, it may be assumed that even when writing about the methods of enhancing patients' treatment, one should still refer to patients as beneficiaries of these methods, not only as those to which these methods apply. As will be demonstrated, the biomedical model has a bearing on the way patients are depicted in medical case reports. Yet, it is not the only factor influencing patient imaging examined in the present study. I will now proceed to describe the data and the methods of my analysis.

3. Data and methods

The corpus for this study comprises fifty case reports taken from four international medical journals aimed at health professionals – *The Lancet* (15), *The Journal of American Medical Association* (12), *The New England Journal of Medicine* (12), and The *British Medical Journal* (11). The issues examined were published between 1995 and 2008 and were devoted to a variety of medical fields. Taavitsainen and Pahta (2000: 60) define the genre of case report in the following way: "In its typical form, the case report records the course of a patient's disease from the onset of symptoms to the outcome, usually either recovery or death. The background and a commentary on the disease are also given, but their scope may vary. Often a limited review of the literature is added and the number of known cases stated".

Generally, case reports present new diseases or diseases that are already known but which have unusual manifestations. The rationale behind the choice of the genre of case report was the fact that this text-type does not present general medical knowledge or detailed results of clinical research, but it discusses particular patients suffering from particular diseases. In other words, case reports give the account of diagnosis and treatment but always in a real context referring to a given person. What is more, it is the genre which brings together all the elements of the medical management of a patient – recounting his/her history and performing physical examination/tests as well as therapy and the results of treatment which are described in respective sections of genre texts. Thus, all the past and present aspects of the patient's health are composed into a report of a case.

In the analysis, each report was carefully examined for words that referred to the patients, but only the descriptions of diagnosis and treatment

were included, leaving out the fragments which concerned demographic information about patients. Then, the sentences containing references to patients were isolated by means of *Wordsmith 5* and examined further.

4. Results and discussion

The present analysis consists of two stages. In the first stage (sub-sections 4.1. - 4.3), patient imaging is examined at the sentence level and with respect to grammar. It has been established that the authors of the texts under study use a variety of techniques in order to distance themselves from the subject of their medical inquiry. To this end, they select specific grammatical configurations which enable them to place words referring to patients in various sentential positions, which, in turn, may affect patients' prominence in texts.

4.1 The removal of agency in the process of diagnosis and treatment

The study reveals that the medical texts under analysis abound in sentences in which diagnostic procedures and treatment are described with no trace of an agent who performs these actions.

(1) He was stabilised with 7.5 mg zopiclone nightly and 3.75 mg daily and monitored closely. B2

(2) He was treated with broadspectrum antibacterial agents (i.e., vancomycin, ceftriaxone, and metronidazole) and antivirals (i.e., acyclovir and foscarnet). JA2

The research on the use of impersonal constructions in scientific discourse indicates that their aim is to focus on what is being studied (Bazerman 1988; Potter 1996; Marco 2000). On the basis of the history of the development of medical practice, it may also be assumed that as the specialisation of medicine has progressed and the methods of treatment have improved, illness (Dubertret 2006: 75) and various medical procedures rather than people who perform these procedures have become more emphasized in texts (Ashcroft 2000: 288). The Passive Voice, as used in (1) and (2), seems to serve its purpose, i.e. drawing attention to medical facts and treatment and hiding the "humble servants of the discipline" (Hyland 2001: 209).

4.2 Patients as locations

In a number of sentences patients do not occur either in the subject or object positions but in prepositional phrases in which the treated are presented as locations of infections and illness. Therefore, rather than to patients themselves, readers' attention is drawn to the diseases examined (Dubertret 2006: 75) and the treatment performed (Ashcroft 2000: 288).

(3) This report describes a case of imported CRS diagnosed **in an infant girl** aged 10 weeks born in New Hampshire to Liberian refugee parents. JA9

(4) Illness **in the two Louisiana residents** was attributed to shellfish that was not prepared or handled properly, perhaps because of difficult living conditions after the hurricanes. LA5

This particular linguistic phenomenon is also a textual manifestation of the container metaphor. Introduced by Lakoff and Johnson in 1980, it presents objects or concepts as having an inside and outside, and as being capable of holding something. As Lakoff and Johnson (1980) explain, "[w]e are physical beings, bounded and set off from the rest of the world by the surface of our skins, and we experience the rest of the world as outside us. Each of us is a container, with a bounding surface and an in-out orientation" (Lakoff and Johnson 1980: 29). When the container metaphor is used in medical discourse, the source domain CONTAINER is mapped onto a patient who "contains" a disease. The concept of disease in a patient is utilised to describe a number of medical procedures or to give account of medical facts. On the one hand, the patient's body tends to be viewed as a container in which diseases are localised and particular treatments performed. The skin, on the other hand, is the "bounding surface" which is crossed by a doctor in an attempt to get from the outside to the inside and manage the diseased site. Language-wise, the container effect is achieved by placing the words referring to patients in the position of complements of prepositional phrases which together function as adverbials of place, here with the meaning of a container. As a result, they serve as the background modifying the topic of a sentence. What comes to the fore is either an instance of a disease that has been localised in a patient (cf. Hodgkin 1985) or a specific treatment that is performed.

4.3 Deverbal nouns and derived nominals

In some other sentences, medical treatment and procedures are presented as the focal points. These often occupy the subject position and may be realised in the form of deverbal nouns (5 - 6) or derived nominals (7).

(5) **His clinical evaluations** did not suggest severe illness, and two bacterial throat cultures were negative. JA2

(6) **Examination** of the ear, nose, and throat detected no discharge or signs of inflammation. NEJM9

(7) Guidelines issued by the Public Health Service and the Infectious Disease Society of America recommend **screening all patients** for active tuberculosis and obtaining mycobacterial blood cultures before rifabutin prophylaxis is begun. NEJM3

The two forms enable the writer to remove his/her presence and to draw readers' attention to the described procedure. As regards patients, they are often made visible through the form of possessive pronouns or located in *of*-phrases. These linguistic selections seem to reflect the assumptions of the *biomedical model of medicine*. As has already been mentioned, this model views patients as those who passively undergo treatment. Therefore, what comes to the fore is treatment (in a nominal form) (cf. Taavitsainen – Pahta 2000: 71; Ashcroft 2000: 288).

I shall now proceed to the second stage of the analysis (sub-sections 4.4 - 4.7) in which patient imaging is examined on the level of text and with respect to lexis.

4.4 The separation of biological processes from the person (de-personalization) (Anspach 1988)

When studying case presentations delivered by physicians during ward rounds, Anspach (1988) reported the separation of biological processes from the person (de-personalization)[1]. What Anspach (1988) points to concerns such utterances as those where the main focus falls not on a patient but

[1] Case presentations are highly conventionalized oral descriptions of patients and their diseases, which are performed by clinicians or medical students in clinical settings (cf. Atkinson 1995).

on a disease or an organ. The author acknowledges the fact that doctors know the referents, but people and their treated organs or other body-parts are not rendered integral (1988: 366). The same practices are evident in the medical case reports of the corpus.

(8) Chest radiographs revealed progressive consolidation and a new right pleural effusion (Figure 2B). A chest tube was placed in the left pleural space. Total white blood cell count increased to 43.6 × 103/μL (83% segmented neutrophils, 12% lymphocytes, and 5% monocytes) and serum creatinine level increased to 3.7 mg/dL (327 μmol/L). Hematocrit, platelet count, liver enzymes, and coagulation profile remained normal, with the exception of an aspartate aminotransferase level of 61 U/L. JA3

(9) We started a magnesium infusion to maintain ionised magnesium levels of 1·5–2·0 mmol/L, asmuscle spasms were consistently worse once the serum magnesium fell below 1·5 mmol/L. Painful muscular spasms continued for weeks after extubation and were controlled by supranormal magnesium levels for a further 9 days and subsequently baclofen. LA14

In (8), medical tests are performed and their results are described, yet, with no indication of those who undergo the tests. Interestingly enough, even when "[a] chest tube was placed in the left pleural space", the patient was not mentioned. The second instance deals with administering drugs and giving account of specific symptoms, but again these facts are presented separately from the people receiving the treatment. It needs to be stressed that the de-personalization of patients as reported by Anspach (1988) and in the present study occurs in larger fragments of utterances and in texts, respectively.

Although it may seem difficult to evaluate patients' visibility given only fragments of their contexts, the corpus contains numerous examples in which diseases, processes, symptoms, particular treatments, or even body-parts are presented as if in abstraction from a patient. It is perhaps more understandable not to mention patients when writing about certain detailed analyses of specimens, tests on the cellular level, etc. Yet, mentioning patients, those whom the experience concerns, as well as describing their symptoms or reactions may also be considered requisite. Additionally, it should be stressed that it is not the possible difficulty in identifying who undergoes the

treatment dealt with in the case reports, as, by definition, case reports tend to present one particular patient who suffers from a previously unknown disease or a disease which manifests itself in an unusual way. The issue is whether patients' presence should be given more prominence in those parts of texts which refer to the matters directly affecting the treated and their experience of illness.

4.5 Focus on body-parts

Under the label of *focus on body-parts* are subsumed examples in which the human body-parts, organs, tissues, etc. are the most salient element of sentences (9) and/or are rendered separate from the patient (10-11):

(9) […] **the infant's right ear** passed the screening test but *the left ear* required further evaluation by an audiologist. JA9

(10) **The left eye** had 3+ anterior vitreous cells, an engorged disk, and cystoid macular edema. NEJM5

(11) **The bladder** appeared to be decompressed and contained an indwelling catheter. NEJM13

According to Virchow (1880), whose work on autopsy and pathological anatomy underlies the biomedical model, all diseases stem from the dysfunction of tissues. Following this medical premise, organs and tissues claimed centrality in medical case writing (Nowell-Smith 1995: 52) as they began to be perceived as the location of illness. This fact from the history of medicine might be helpful in explaining the linguistic phenomenon of the subjectivization of body-parts.

4.6 Technology as the agent (Anspach 1988)

A number of studies of scientific discourse have addressed the feature of emphasizing the role of data i.e. it is data that "show", "prove" or "reveal". To describe this characteristic, Potter coined the term "data primacy" (1996: 153). He claims that the fact that the agent is deleted from the text contributes to rendering the information objective and independent from

human involvement (Potter 1996: 153). A similar practice has been found in spoken medical discourse. For instance, one of the features of medical case presentations is "treating technology as the agent" (Anspach 1988), where it is medical equipment and diagnostic procedures that "reveal" or "show" particular results of the study. Such examples have also been found in the corpus under analysis:

(12) **A chest radiograph revealed** a nodular infiltrate, which was thought to be a residual finding from pneumonia diagnosed in early January. JA11

(13) Pituitary MRI was normal; inferior petrosal sinus sampling excluded pituitary-dependent Cushing's disease; and octreotide and **CT scans of chest, abdomen, and pelvis showed** no abnormality other than bulky adrenal glands consistent with adrenal hyperplasia. LA11

The focus on technology constitutes another textual reflection of the development of medicine, i.e. the introduction of modern diagnostic procedures (Ashcroft 2000). These technological innovations render the patient's body readable (cf. Atkinson 1995) and let the data speak for themselves.

4.7 Patients as cases

The analysis of the texts in the corpus also demonstrates the common practice of referring to an individual occurrence of particular disease as a *case*. Yet, close scrutiny also reveals examples in which the word *case* refers not to a disease but to a patient (cf. Fowler 1996: 124-134):

(13) Two of the 18 US **cases** of inhalational anthrax reported prior to the recent bioterrorism-related outbreak had underlying lung disease; one had beryllium exposure and chronic pulmonary fibrosis and the other had underlying pulmonary sarcoidosis. JA3

(14) Many of the reported **cases** are children and only two cases have survived. LA3

These examples suggest treating a patient as "the object of some disease entity" (Mead – Bower 2000: 1089), which, in consequence, can lead to other similar naming practices among health professionals (cf. Anspach 1988).

5. Conclusion

The present analysis of medical case reports reveals that some of their features have the potential to dehumanize patients. On the one hand, the use of specific grammatical structures influences patients' sentential position, which may affect their prominence in texts. On the other hand, the very decision of whether or how to mention a patient when describing his/her illness determines his/her textual visibility. Consequently, patient imaging in the corpus under study seems incomplete; because it renders patients, their body-parts, and biological processes separate. Such a writing practice requires further research as what distinguishes medicine from other scientific areas is that it does not only study diseases and develop new ways of their treatment but also manages patients who suffer all the consequences of being ill. While the alternative models of medicine presented have sensitized physicians to consider this human factor during any form of direct contact with patients (be it consultation, end-of-life conversation, etc.), it seems that similar sensitization could also be executed in the realm of medical publications. When giving account of the innovative techniques of treatment or manifestations of diseases clinicians also mediate a given image of patients, either as objects or subjects of medical study and practice. It appears to be an issue of concern in light of the fact that texts written by the already established members of the profession not only acquaint novices with particular attitudes and values but also promote discipline-specific modes of writing.

REFERENCES

Sources

Barakat, L. A. et al.
2002 "Fatal inhalational anthrax in a 94-year-old Connecticut woman", *Journal of American Medical Association* 287, 863-868. JA3

Bishai, W. R. et al.
1996 "Rifampin-resistant tuberculosis in a patient receiving Rifabutin prophylaxis", *The New England Journal of Medicine* 334, 1573-1576. NEJM3

Bush, Larry M. et al.
2001 "Index case of fatal inhalational anthrax due to bioterrorism in the United States", *The New England Journal of Medicine* 345, 1607-1610. NEJM9

Carroll, D. N. et al.
2005 "Congenital viral infection?", *The Lancet* 365, 1110. LA5
Jones, I. R. – G. Sullivan
1998 "Physical dependence on zopiclone: Case reports", *The British Medical Journal* 316, 117. B2
Keenan, Niall et al.
2006 "Unexpected shortness of breath in a patient with Cushing's syndrome", *The Lancet* 367, 446. LA11
Kellenberg, J. et al.
2005 "Brief report: Imported case of congenital rubella syndrome – New Hampshire, 2005", *Journal of American Medical Association* 295, 492-493. JA9
Kenyon, C. et al.
2004 "Fatal case of unsuspected pertussis diagnosed from a blood culture – Minnesota, 2003", *Journal of American Medical Association* 291, 1557-1558. JA11
Lindley-Jones, M. et al.
2004 "Recurrent tetanus", *The Lancet* 363, 2048. LA14
Mak, R. et al.
2001 "Two fatal cases of adenovirus – related illness in previously healthy young adults – Illinois, 2000", *Journal of American Medical Association* 286, 553-556. JA2
Rickman, L. S. et al.
1995 "Uveitis caused by tropheryma whippelii (Whipple's Bacillus)", *The New England Journal of Medicine* 332, 363-366. NEJM5
Sayegh, Mohamed H. – Robert B. Colvin
2003 "Case 8-2003: A 35-year-old man with early dysfunction of a second renal transplant", *The New England Journal of Medicine* 348, 1033-1045. NEJM13
White, J. M. et al.
2004 "Granulomatous amoebic encephalitis", *The Lancet* 364, 220. LA3

Special studies

Albert, T.
2004 "Why are medical journals so badly written?", *Medical Education* 38, 6-8.
Anspach, R. R.
1988 "Notes on the sociology of medical discourse: The language of case presentation", *Journal of Health and Social Behaviour* 29, 357–375.
Ashcroft, R. E.
2000 "Teaching for patient-centred ethics", *Medicine, Health Care and Philosophy* 3, 287-295.

Atkinson, D.
2001 "Scientific discourse across history: A combined multi - dimensional/ rhetorical analysis of the Philosophical Transactions of the Royal Society of London". In: S. Conrad and D. Biber (eds.) *Variation in English: Multi-dimensional Studies.* Harlow: Longman, 337-374.
Atkinson, P.
1995 *Medical Talk and Medical Work*. London: Sage Publications.
Bazerman, C.
1988 *Shaping Written Knowledge. The Genre and Activity of the Experimental Article in Science*. Madison: The University of Wisconsin Press.
Beagan, B. L.
2000 "Neutralizing differences: Producing neutral doctors for (almost) neutral patients", *Social Science & Medicine* 51, 1253-1265.
Dubertret, L.
2006 "Patient – based medicine", *Journal of European Academy of Dermatology and Venerology* 20, 73-76.
Engel, G. L.
1977 "The need of a new medical model: A challenge for biomedicine", *Science* 196, 129-136.
Fowler, R.
1996 *Language in the News: Discourse and Ideology in the Press*. London: Routledge.
Freidson, E.
1970 *Profession of Medicine. A study of the Sociology of Applied Knowledge*. New York: Harper & Row.
Gotti, M. – F. Salager-Meyer (eds.)
2006 *Advances in Medical Discourse Analysis – Oral and Written Contexts*. Bern: Peter Lang.
Hodgkin, P.
1985 "Medicine is war", *British Medical Journal* 291, 1820-1821.
Hyland, K.
2001 "Humble servants of the discipline? Self - mention in research articles", *English for Specific Purposes* 20, 207-226.
Kenny, N. P. – B. L. Beagan
2004 "The patient as text – a challenge for problem – based learning", *Medical Education* 38, 1071-1079.
KIAP – Cultural Identity in Academic Prose: National vs. discipline – specific.
2006 http://www.uib.no/kiap/index-e.htm.
Lakoff, G. – M. Johnson
1980 *Metaphors We Live by*. Chicago: The University of Chicago Press.
Lock, M. – D. Gordon (eds.)
1988 *Biomedicine Examined*. Dordrecht: Kluwer Academic Publishers.
Marco, M. J. L.
2000 "Collocational frameworks in medical research papers: A genre based study", *English for Specific Purposes* 19, 63-86.

Mead, N. – P. Bower
2000 "Patient – centredness: A conceptual framework and review of the empirical literature", *Social Science & Medicine* 51, 1087-1110.
Myers, G.
1990 *Writing Biology: The Social Construction of Popular Science: The Narrative of Science and the Narrative of Nature*. Madison: University of Wisconsin Press.
Nowell-Smith, H.
1995 "Nineteenth - century narrative case histories: An inquiry into stylistics and history", *CBMH/BCHM* 12, 47-67.
Potter, J.
1996 *Representing Reality : Discourse, Rhetoric and Social Construction*. London: Sage.
Sontag, S.
1991 *Illness as Metaphor and AIDS and Its Metaphors*. London: Penguin.
Taavitsainen, I. – P. Pahta
2000 "Conventions of professional writing: The medical case report in a historical perspective", *Journal of English Linguistics* 28, 60-76.
Virchow, R.
1880 *Post-Mortem Examinations, with Special Reference to Medico-Legal Practice*. Translated by T. Smith. Philadelphia: Presley Blackiston.
Wade, D. T. – P. W. Halligan
2004 "Do biomedical models of illness make for good healthcare systems?", *British Medical Journal* 329, 1398-1401.

A corpus-based study of modal adverbs in English from the viewpoint of grammaticalization*

Daisuke Suzuki
Kyoto University

ABSTRACT

This study sheds light on modal adverbs in present-day English with respect to certain processes of linguistic change, paying particular attention to two groups of synonymic expressions: *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely*. After extracting data regarding these adverbs from the British National Corpus, the study aims to determine two factors regarding their patterns of occurrences: (i) whether they occur in initial, medial or final position, and (ii) which pronouns fill the Subject slot in their clauses. The results of this analysis show that the modal adverbs differ in the part that they play in fulfilling the communicative function; that is, at a discourse-pragmatic level. Moreover, I suggest that *no doubt* and *surely* are further advanced than the others in the processes of intersubjectification.

1. Introduction

The synonymic expressions *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely* are nearly equivalent in meaning to the others of each group, and are thus classified in the same semantic category. As demonstrated by Examples (1a-c) and (2a-c), these six adverbs are used to express a speaker's judgment regarding the certainty or probability of a proposition:

* This article is a revised and extended version of Suzuki (2011). I would like to express my sincere gratitude to Yoko Iyeiri, Masa-aki Tatsuki, Kensei Sugayama, and Masa-aki Yamanashi for their helpful suggestions on my research. I would also like to thank the anonymous reviewers for their helpful comments on earlier versions of this paper. This research was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (No. 22·3983). All remaining errors and inadequacies are, of course, my own.

(1) a. It was *no doubt* clever of him to offer his resignation at that point in the proceedings. (Quirk et al. 1985: 622)
b. You have *doubtless* or *no doubt* heard the news. (Fowler 1998: 230)
c. During the action the person will *undoubtedly* have certain feelings towards it and gain satisfaction from achievement. (ACAD) (Biber et al. 1999: 854)

(2) a. It will *certainly* rain this evening. (Swan 2005: 20)
b. He has *surely* made a mistake. (Huddleston – Pullum 2002: 767)
c. Ruth was *definitely* at Goosehill School. (CONV) (Biber et al. 1999: 853)

Despite their similarity in form and meaning, as Biber et al. (1998) mention, investigating the use and distribution of synonyms in a corpus allows us to determine their contextual preferences. To conduct the detailed analysis required to distinguish among them, I investigated corpus data to consider the functions and patterns of usage of each adverb. Thus, I could identify the factors significant in predicting each adverb's usage and the way some differs from the other adverbs.

2. Previous studies

Before commencing the actual study, this section examines the previous findings regarding the usage of *no doubt*, *doubtless*, and *undoubtedly*. Based on their research, Huddleston – Pullum (2002: 768) propose that modal adverbs express one of four levels of strength according to the speaker's commitment: (i) strong, (ii) quasi-strong, (iii) medium, and (iv) weak. They place *undoubtedly* into the strong category and *doubtless* into the quasi-strong category. Biber et al. (1999) classify *no doubt* and *undoubtedly* as doubt and certainty adverbials, which show the speaker's certainty or doubt about propositions. In addition, Simon-Vandenbergen – Aijmer (2007: 134) indicate that in terms of the degree of probability, the meaning of *undoubtedly* is in-between *certainly* and *no doubt*, and in terms of subjectivity, *undoubtedly* seems to have a more objective reading than either *certainly* or *no doubt*.

However, Quirk et al. (1985: 623) mention that similar to *no doubt*, *doubtless* implies some doubt and is synonymous with "very probably," while *undoubtedly* expresses conviction. In addition, Konishi (2006: 420) points out that *no doubt* and *doubtless* convey nearly the same meaning, whereas *no doubt* is a little stronger in meaning than *doubtless*.

Focusing on *no doubt*, Simon-Vandenbergen and Aijmer (2007) reveal the semantics and pragmatics of *no doubt* in various contexts such as expressing a high degree of probability and concessive meaning, and its ability to function as a discourse marker as illustrated in Examples (3a, b):

(3) a. *No doubt*, money played its part in this (ICE-GB:W2C-007/64)
b. Britain and Germany will *no doubt* continue to disagree on particular policy issues *but* Chancellor Kohl and John Major clearly feel that the important thing is to have the kind of ongoing relationship … (ICE-GB:S2B-002/105)

Moreover, they show the possibility of the following development of *no doubt* from the viewpoint of grammaticalization:

(4)

Existential	>	*no doubt about it*	>	*no doubt*
+ certain		+ certain		+ probable
+ objective		± subjective		+ subjective

(Simon-Vandenbergen – Aijmer 2007: 127)

Turning to the usage of *certainly*, *surely*, and *definitely*, which are all classified into the strong category in Huddleston – Pullum's (2002: 768) classification, the works of Hoye (1997) and Simon-Vandenbergen – Aijmer (2007) reveal some syntagmatic behaviours of the adverbs. Hoye (1997) indicates that *certainly* collocates with all modals, although restrictions apply in the case of combinations with epistemic modals of possibility, as shown by Example (5a). It is also pointed out that *definitely* remains too lexically loaded to occur in the environment of epistemic *must*, where *certainly* may be a more natural choice, as in Example (5b), and it is probably more emphatic in non-epistemic contexts, as in Example (5c):

(5) a. *He *certainly* might/may be there. (Hoye 1997: 162)
b. His behavior must *certainly*/*?definitely* bring him to grief sooner or later. (Ibid.: 163)
c. I *definitely* won't do it and let that be an end to the discussion! (Ibid.: 163)

Focusing on *surely*, Hoye (1997: 191) states that "regardless of its syntactic position, *surely* functions to seek agreement in anticipation of some opposition

and is not purely used for the reinforcement of truth-value; it also tends to precede a question"[1]. Examples (6a, b) illustrate this point:

(6) a. *Surely* they couldn't have expected you to complete the project so soon (could they)? (Ibid.: 191)
b. *Surely* the run must be nearly over now. (W. 11.2.203)

In this respect, Simon-Vandenbergen – Aijmer (2007: 119) maintain that "*certainly* expresses certainty based on the speaker's subjective assessment, while *definitely* expresses certainty based on the permanent nature of a state of affairs". With regard to *surely*, they also argue that its functions depend on its position in the clause as well as the context. According to them, when *surely* is in initial position, where it frequently collocates with a modal verb, the speaker very clearly expresses an opinion on what is possible/likely or what is desirable (pp.135-137). On the other hand, when it is in medial position, *surely* is considered to lose its epistemic meaning and functions as an emphasizer or intensifier (pp.138-139).

Although the existing literature interprets variously the employment of the two groups of synonymic expressions *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely*, it offers no clear-cut usage determinant for them. To fill this research gap, this paper proposes some new, clear guidelines for their use from the viewpoint of intersubjectification. The following sections embark upon a further interpretation of the functions of the six modal adverbs, based on the analysis of a large body of language.

3. Data and method

This corpus-based investigation offers the total numbers of occurrences of the two groups of synonymic modal adverbs *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely*, which enable us to use quantitative analysis. As the source of the data for analysis, I selected the British National Corpus (BNC) because of its large scale and wide range, and thus ability to provide many instances of the use of the six modal adverbs for various purposes within various contexts[2]. To prepare the data for analysis, I first extracted all

1 A similar view is expressed by Aijmer (2009) and Downing (2001), who consider *surely* as mainly challenge and seeking agreement.

2 I made use of Mark Davies' freely available on-line interface (http://corpus.byu.edu/bnc/).

occurrences of the adverbs from the corpus and identified 5,955 occurrences of *no doubt*, 844 of *doubtless*, and 2,343 of *undoubtedly*; 18,118 of *certainly*, 6,032 of *surely*, and 3,056 of *definitely*. I then examined each occurrence to identify those in which one of the six adverbs functioned as a sentence adverb[3] and identified **2,701** instances of *no doubt*, **731** of *doubtless*, **2,202** of *undoubtedly*; **15,718** of *certainly*, **5,369** of *surely*, and **2,350** of *definitely*. I further conducted quantitative analysis on these in terms of frequency as described in the following sections.

In my analysis of *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely*, I focused attention on information provided by the larger context in which those expressions occur, so my primary consideration was investigating the two factors regarding their patterns of occurrence: (i) whether they occur in initial, medial or final position[4], and (ii) which pronouns fill the Subject slot in their clauses. I therefore first determined the number of instances in which each adverb occurred in each position, as well as the number of instances in which it occurred with pronouns, to calculate the frequency of its occurrence in terms of positioning and function. In order to illuminate their functions in detail, I then examined the percentage of occurrences of these adverbs with each of the pronouns in the BNC.

In this study, I attempt to account for the developments of *no doubt* and *surely* as pragmatic markers in terms of (inter)subjectification as in the following:

(7) a. (clause-internal) adverb > conjunction/sentential adverb > pragmatic marker (Brinton 2008: 27)

---

3 For this analysis, I excluded all the examples of such idiomatic phrases as "slowly but surely" regarding those of *surely*, and excluded those of the modal adverbs modifying not a clause but a phrase in which a comma (,) intensifies the expressed meaning, as in the following:

i Given the political will, a primitive nationalism can be generated by governments in a remarkably short space of time, *certainly* in less than a generation. (AE8)

ii …, and that a group of glaucous and ivory gulls were standing around, *no doubt* waiting in the hope of leftovers. (CRJ)

4 In Hoye (1997) and Quirk et al. (1985), the positions in which they appear are presented as follows:

I	(*initial*)	*Possibly* they may have been sent to London.
iM	(*initial-medial*)	They *possibly* may have been sent to London.
M	(*medial*)	They may *possibly* have been sent to London.
mM	(*medial-medial*)	They may have *possibly* been sent to London.
eM	(*end-medial*)	They may have been *possibly* sent to London.
iE	(*initial-end*)	They may have been sent *possibly* to London.
E	(*end*)	They may have been sent to London *possibly*. (Hoye 1997: 148)

b. non-/less subjective > subjective > intersubjective (Traugott 2010: 35; Traugott and Dasher 2002: 225)

In characterizing (inter)subjectivity, we see several notions of subjectivity in cognitive and functional linguistic literature, including Lyons' (1977), Traugott's (1989, 2010), Langacker's (1990, 1999), and Nuyts' (2001, 2012). Focusing on functions of language and its semantic change, I draw on Traugott's stand, which involves synchronic as well as diachronic clines, so subjective meanings can also be intersubjective on the basis of the ambient context.

4. Results and discussion

In this section, I describe the means by which I elucidated the functions of *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely* from the perspective of discourse and conversation. First, I determined the frequency with which each adverb occurs in terms of positioning. As shown in Table 1, Biber et al. (1999: 872) identify a tendency for stance adverbials to be positioned in medial positions[5]:

Table 1. Positioning of stance adverbials across registers (Based on Biber et al. [1999: 872])

	Initial position (%)	Medial position (%)	Final position (%)
CONVERSATION	•••	••••••••••	•••••••
FICTION	•••••	••••••••••••	••••
NEWSPAPER	•••••••	••••••••••••	••
ACADEMIC	••••••	••••••••••••••	•

each • represents 5%

Figs. 1 and 2 illustrate the percentage of total occurrences in which the six adverbs are positioned in the initial, medial, and final position[6], with such

5 According to Biber et al. (1999), stance adverbials can be classified into three categories: epistemic, attitude, and style adverbials. Epistemic adverbials contain *probably*, *I think*, *in fact*, *really*, *according to*, *mainly*, *generally*, *in my opinion*, *kind of*, *so to speak*, in addition to *no doubt*, *undoubtedly*, *certainly*, and *definitely*; attitude adverbials include *unfortunately*, *to my surprise*, and *hopefully*; and style adverbials include *frankly*, *honestly*, *truthfully*, and *in short*.

6 The data of Figs. 1 and 2 are given in the appendix.

positioning linguistically illustrated in Examples (8a-i) and (9a-i). Despite the tendency for stance adverbials to be positioned in medial positions, we can observe that *no doubt* and *surely* are positioned in the clause-initial position relatively frequently in the BNC. As Halliday – Matthiessen (2004), Hoye (1997), and Halliday (1970) observe, a modal adverb positioned initially expresses the topic or theme of modality[7], so that there is a strong tendency for *no doubt* and *surely* to function as markers of topics in discourse.

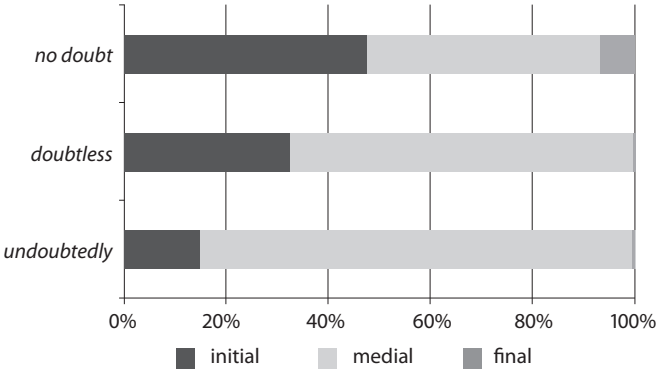


Figure 1. Proportions of the positions of *no doubt*, *doubtless*, and *undoubtedly* (BNC)

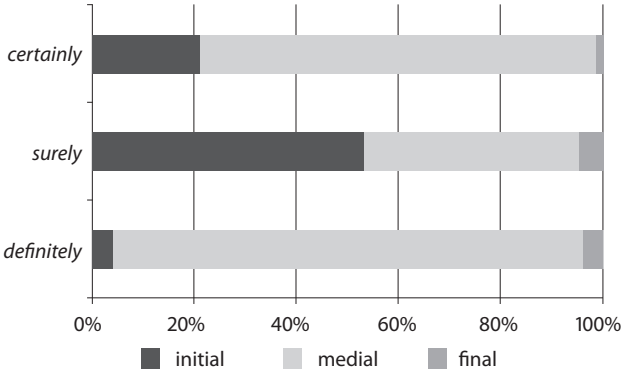


Figure 2. Proportions of the positions of *certainly*, *surely*, and *definitely* (BNC)

7 As Halliday – Matthiessen (2004), Hoye (1997), and Halliday (1970) argue, Examples (i, ii) convey the same meaning in terms of probability, but the use of *possibly* in Example (i) serves the discourse function of expressing the topic or theme.

i **Possibly** it was Wren.

ii It **may** have been Wren. (Halliday 1970: 335)

Table 2. Frequencies of *no doubt*, *doubtless*, and *undoubtedly* with clause subject pronouns (BNC)[8]

	no doubt		*doubtless*		*undoubtedly*	
	Freq.	Per 1,000	Freq.	Per 1,000	Freq.	Per 1,000
I	49	18.14	9	12.31	9	4.09
you	182	**67.38**	27	36.94	29	13.17
he	209	77.38	49	67.03	90	40.87
she	88	32.58	11	15.05	21	9.54
it	187	69.23	52	71.14	182	82.65
we	55	20.36	4	5.47	13	5.90
they	120	44.43	30	41.04	72	32.70
this	92	34.06	18	24.62	92	41.78
that	28	10.37	9	12.31	24	10.90
these	3	1.11	2	2.74	5	2.27
those	1	0.37	0	0.00	0	0.00

Table 3. Frequencies of *certainly*, *surely*, and *definitely* with clause subject pronouns (BNC)

	certainly		*surely*		*definitely*	
	Freq.	Per 1,000	Freq.	Per 1,000	Freq.	Per 1,000
I	1.413	89.9	110	20.5	272	115.7
you	517	32.9	412	**76.7**	136	57.9
he	928	59.0	303	56.4	139	59.1
she	388	24.7	167	31.1	99	42.1
it	2.531	161.0	672	125.2	270	114.9
we	671	42.7	198	36.9	115	48.9
they	771	49.1	224	41.7	115	48.9
this	380	24.2	245	45.6	57	24.3
that	336	21.4	255	47.5	93	39.6
these	36	2.3	9	1.7	12	5.1
those	13	0.8	2	0.4	1	0.4

8 In Tables 2 and 3, the raw frequencies of pronouns are given for each form, followed by the normalized figure of the number of occurrences per 1,000 instances.

(8) a. *No doubt* some of them volunteered for war service a year later. (B1P)
b. Quarterback Stan Gelbaugh was *no doubt* harbouring similar thoughts. (AKE)
c. We shall be a bit cold for that reason *no doubt*. (KCN)
d. *Doubtless* he was trying to convince himself as much as us. (FS0)
e. The ascent was *doubtless* relatively easy. (EFR)
f. Other factors contributed to the Tory revival, Scottish hard-headedness about matters of the pocket *doubtless* among them. (AK9)
g. *Undoubtedly* the product that set the pace was Aldus PageMaker. (G00)
h. The primary cause was *undoubtedly* a thermal plume. (CKC)
i. It had its effect on our lives *undoubtedly*, particularly mine. (C8T)

(9) a. *Certainly* he has never given a hint that he knows anything. (ASN)
b. He's *certainly* got something to shout about. (K23)
c. That's usual, *certainly*. (G3E)
d. *Surely* there are more testing comparisons to be made. (FS8)
e. Egypt is *surely* the motif of the year. (G06)
f. That sounds odd though, *surely*. (KPV)
g. *Definitely* I have some ideas. (CK4)
h. He is *definitely* in contention. (CH7)
i. So it's a mapping *definitely*. (GYX)

To support this analysis of *no doubt* and *surely* in interpersonal contexts explicitly, the quantitative distribution of the instances of the six adverbials among the clause subjects is presented in Tables 2 and 3. It is shown that *no doubt* and *surely* display a tendency toward a higher frequency of co-occurrence with clause subject pronoun *you*, in contrast with the other four expressions: *doubtless*, *undoubtedly*, *certainly*, and *definitely*.

This point is linguistically illustrated in Examples (10a, b) and (11a, b), in which as clearly shown by co-occurrence with second-person pronouns (and in interrogatives), we can observe the intersubjective uses of *no doubt* and *surely* occur:

(10) a. "It means that any office development or commercial of this kind in this I twelve policy that was to occur in Harrogate would not be

counted off the sixty hectares of em erm of I five allocation, it would be in addition to it, and that may be a very important breakthrough for us. Well yes. *No doubt* **you**'ll want to return to that. Thank you very much." (JAD)

b. "… should it ever become necessary to activate this device, I assure you that both I and my staff will be well clear of its effects. *No doubt* **you** are disappointed to hear that." "You're right, Nate." (HJD)

(11) a. Not so very. Not so very. You know her. I think you know her. Anyway you know all about her, for I have told you. And you have seen her. *Surely* **you** remember her? They say she is very clever and will keep my house well. This is more important than a long nose, don't you think? (CDX)

b. Obviously, you're fond of him, but *surely* **you** can see? His accent, Edouard. You'll never eliminate it altogether, you know. (C8S)

Moreover, the marked pattern of the use in interrogative contexts can be observed in the BNC. The following are examples of *no doubt* and *surely* used as metalinguistic devices to confirm or emphasize information and understanding between the speaker and hearer; that is, to fulfill interpersonal functions:

(12) a. **You** have heard different versions, *no doubt*? (G1A)
b. And then **you** called in the Royal Oxford, *no doubt*? (HWM)
c. **You** have read it, *no doubt*? (GVP)
d. **You** have heard of Vechey's death, *no doubt*? (H98)

(13) a. **You** don't have to eat the skin, *surely*? (KBW)
b. **You** can wait until then, *surely*? (EVC)
c. **You** wouldn't destroy it, *surely*? (HNJ)
d. Oh **you** can grate the cheese *surely*? (KD5)

In a nutshell, *no doubt* and *surely* evolve from modal adverbs to pragmatic markers, and come to serve interpersonal functions. Thus, the two adverbs have undergone the changes identified with intersubjectification, as shown in Table 4:

Table 4. Summary of the six modal adverbs

	no doubt, surely	
doubtless, undoubtedly, certainly, definitely		
adverb	>	**pragmatic marker**
epistemic	>	**interpersonal**
subjective	>	**intersubjective**
	intersubjectification	

5. Conclusion

This study examined the modal sub-systems of *no doubt*, *doubtless*, *undoubtedly* and *certainly*, *surely*, *definitely* from a functional perspective. In particular, examining the functions of the six adverbs from the viewpoints of discourse and conversation, I demonstrated that *no doubt* and *surely* differ in the parts that they play at the discourse-pragmatic level.

I also elucidated that intersubjectification; that is, focus on the addressee, motivates the semantic changes affecting no doubt and surely. It follows that the evolutions of the modal adverbs are best understood as the processes of intersubjectification, and that Traugott's stand on (inter) subjectification is generally valid in analyzing the use of modal adverbs, such as those examined in this study.

REFERENCES

Sources

British National Corpus (BYU-BNC)
Maintained by Mark Davis (Brigham Young University).
http://corpus.byu.edu/bnc/.

Special studies

Aijmer, Karin
2009 "The pragmatics of adverbs". In: G. Rohdenburg – J. Schlüter (eds.) *One Language, Two Grammars?: Differences between British and American English*. Cambridge: Cambridge University Press, 324-340.

Biber, Douglas et al.
1998 *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.
1999 *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
Brinton, L. J.
2008 *The Comment Clause in English: Syntactic Origins and Pragmatic Development.* Cambridge: Cambridge University Press.
Downing, Angela
2001 "'Surely you knew!' *Surely* as a marker of evidentiality and stance", *Functions of Language* 8 (2), 253-285.
Fowler, H. W.
1998 *Fowler's Modern English Usage* (3rd edn.). Revised by R. W. Burchfield. Oxford: Oxford University Press.
Halliday, M. A. K.
1970 "Functional diversity in language as seen from a consideration of modality and mood in English", *Foundations of Language* 6, 322-361.
Halliday, M. A. K. – Christian M. I. M. Matthiessen
2004 *An Introduction to Functional Grammar* (3rd edn.). London: Arnold.
Hoye, Leo
1997 *Adverbs and Modality in English*. London: Longman.
Huddleston, Rodney – Geoffrey K. Pullum
2002 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Konishi, Tomoshichi (ed.)
2006 *Gendai Eigo Gogen Ziten* [Dictionary of Present-day English Usage]. Tokyo: Sanseido. [In Japanese]
Langacker, R. W.
1990 "Subjectification", *Cognitive Linguistics* 1, 5-38.
1999 *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.
Lyons, John
1977 *Semantics*, Vol. 2. Cambridge: Cambridge University Press.
Nuyts, Jan
2001 *Epistemic Modality, Language, and Conceptualization*. Amsterdam: John Benjamins.
2012 "Notions of (inter)subjectivity", *English Text Construction* 5 (1), 53-76.
Quirk, Randolph et al. (eds.)
1985 *A Comprehensive Grammar of the English Language*. London: Longman.
Simon-Vandenbergen, Anne-Marie – Karin Aijmer
2007 *The Semantic Field of Modal Certainty: A Corpus-Based Study of English Adverbs*. Berlin: Mouton de Gruyter.
Suzuki, Daisuke
2011 "On the Pragmatic Functions of Modal Adverbs: A Corpus-based Study of Linguistic Variations in Present-day English", *Zephyr* 24, 1-21.

Swan, Michael
2005 *Practical English Usage* (3rd edn.). Oxford: Oxford University Press.
Traugott, E. C.
1989 "On the rise of epistemic meanings in English: An example of subjectification in semantic change", *Language* 65, 31-55.
2010 "(Inter)subjectivity and (inter)subjectification". In: K. Davidse et al. (eds.) *Subjectification, Intersubjectification and Grammaticalization*. Berlin: De Gruyter Mouton, 29-71.
Traugott, E. C. – Richard B. Dasher
2002 *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

APPENDIX

Data for Figs. 1 and 2

Form	Initial	Medial	Final	Total
no doubt	1,288	1,237	176	2,701
doubtless	237	492	2	731
undoubtedly	325	1,873	4	2,202

Form	Initial	Medial	Final	Total
certainly	3,261	12,218	239	15,718
surely	2,863	2,275	231	5,369
definitely	89	2,183	78	2,350

The Tübingen Corpus of Eastern European English (TCEEE): From a small-scale corpus study to a newly emerging non-native English variety

Elena Salakhyan

Eberhard Karls University of Tübingen

ABSTRACT

Research in the field of World Englishes aims to pin down, as precisely as possible, the linguistic and pragmatic properties a certain variety displays or does not display. The status of English in the Expanding Circle has been of significant interest in recent years (Berns 1995, 2005; House 2002; Knapp – Meierkord 2002; Jenkins 2007; Sedlhofer – Widdowson 2009, etc.). Nevertheless, the use of English by Slavic speakers in Post-Soviet Space has been largely ignored. Given the typological similarities among the Slavic languages (and similar historical and societal developments in the region) the paper proposes to view the Eastern European English(es) as a variety of English within the Expanding Circle. In particular, the paper questions which morphosemantic patterns, especially those of tense and aspect, emerge in the data. The study draws on spontaneously produced language data of fifteen Slavic speakers of English with L1 Ukrainian, Russian, Polish or Slovak which have been compiled into the Tübingen Corpus of Eastern European English (TCEEE: sixty thousand words). The paper argues that a variety of Eastern European English(es) is indeed emerging and that further studies examining the domains of morphosyntax, morphosemantics and lexis are necessary to provide additional evidence of this development.

1. Introduction

Within the domain of World Englishes, various scholars have closely examined and described regional non-native varieties of English. The availability of typological feature analyses has in turn led to the need to evaluate and give a particular status to these newly emerging varieties. Indian English

(Mukherjee 2010), Sri Lankan Englishes (Mendis – Rambukwella 2010: 181), East and West African Englishes (Simo Bobda 2000: 185), Malay and Singapore English (Lim – Low 2005: 64), East Asian Englishes (Moody 2007: 209; Takeshita 2010: 265) and Chinese English (Bolton 2003), to name a few, have thus obtained the status of distinct varieties of English. With regard to English in Europe, even though Europe is considered to be one geo-political entity, "Slavic Englishes seem to warrant separate treatment" (Seidlhofer 2010: 355).

Thus, as the use of English by Eastern European and Russian speakers in the ELF context has not been given sufficient attention (for some studies on Slavic English see Ustinova 2006; Proshina – Ettkin 2005; Proshina 2010; Salakhyan 2012), I decided to explore this newly emerging Expanding Circle[1] variety and describe its morphosemantic features.

Slavic languages, especially East Slavic (Russian, Ukrainian and Belarusian) and West Slavic (Polish and Slovak) share many similar linguistic features. Rich morphology, which is often fusional, free word order, rich agreement systems, and the category of aspect are salient features of Slavic languages (Comrie – Corbett 2002: 6-11).

The motivation behind this paper is the need to examine a newly emerging Expanding Circle variety of Eastern European English in order to add to accumulated knowledge about World Englishes.

2. Tense and aspect in the Eastern European English variety

In what follows, I discuss the linguistic features of the newly emerging Eastern European English variety. In this paper, the Eastern European English feature profile will be restricted to the discussion of temporal-aspectual features which, in my opinion, stand out in this Expanding Circle variety of English.

In a typical ESL classroom as well as in a natural second language acquisition environment, the acquisition of temporal and aspectual markers and their subsequent use in spoken narration tends to be particularly

---

1 Kachru's (1988) Circles Model of World Englishes conceptualizes the use of English in terms of three Circles - the Inner Circle (L1 English), the Outer Circle (ESL) and the Expanding Circle (EFL) where the Inner Circle English varieties are 'norm-providing'. Schneider's (2003) model is not based on geography and politics; instead it is based on the underlying processes of language change.

challenging for language learners and consolidated language users. In multifaceted language and culture contact situations in which English acts as a *lingua franca* and speakers are multilingual, the situation becomes even more complex.

Given that the English and Slavic-like temporal-aspectual systems differ immensely, it is necessary to examine how Slavic speakers mark temporal relations and express aspectual perspectives when they are involved in language contact situations. Based on the spontaneously produced spoken data[2], the way in which Slavic speakers render temporality in spoken narration will thus be discussed.

Before we look at the manifestations of tense and aspect in the speakers' performance, let me give a brief overview of what the traditional account of Slavic aspect is based on.

Contrary to English, where the category of aspect is grammaticalized, in Slavic aspectual systems the category of aspect is lexicalized; in other words, all verbs exist in aspectual pairs – perfective and imperfective, e.g. *to read* (imperfective) and *to have read* (perfective) (Dahl 1985). Thus, narrating events that occurred in the past, the Slavic speakers are obliged to decide whether they view the event as perfective (i.e. complete) or imperfective (i.e. incomplete)[3]. Aspect usually combines with tense, to provide the basic structure of the narrative (Bogdan – Sullivan 2009: 50). The tense-aspect form of the verb then tells which function the clause performs in the overall narrative (Bogdan – Sullivan 2009: 50). In contrast to English, where marking past events as progressive or non-progressive is possible, in Slavic languages it is not. In other words, narrating events which occurred in the past, a Slavic speaker is only able to convey that (i) the event occurred in the past, and (ii) the event was either complete or incomplete. Thus, no information as to the process of an action, i.e. progressive vs. non-progressive, as it is in English, is conveyed[4].

---

2 The Tübingen Corpus of Eastern European English (TCEEE) is a small-sized corpus (sixty thousand words) of semi-structured video interviews with fifteen Slavic speakers of English (the speakers' proficiency varies from the B1 to the C1 level according to Common European Framework of Reference for Languages classification), with L1 Russian, Ukrainian, Polish or Slovak. Interview questions elicited introspective data (information about each speaker's English learning history, professional life and involvement in international projects), and the spontaneous production data.

3 The Hopper (1977), Lamb (1991) and Bogdan – Sullivan (2009) studies on Russian and Polish demonstrate the insufficiency of the binary assumption.

4 Bogdan – Sullivan (2009: 43-44) argue, taking Polish into account, that aspect in Slavic languages communicates both aspect (by prefixes or suffixes related to the verb stem) and Aktionsart (from the point of view of the verb stem).

The differences between the Slavic-like and the non-Slavic-like temporal-aspectual systems lie not only in the overt manifestations of tense and aspect but also in how Slavic and non-Slavic speakers view reality and the representation of past and non-past events.

An account proposed by Durst-Anderson (1994) attempts to link overt linguistic manifestations with cognition and mental models (Durst-Andersen 1994: 81). As he claims, events and processes first find their manifestation in human cognition and only then are realized in a real language (Durst-Andersen 1994: 62). The central claim of Andersen's argument thus revolves around the assumption that human cognition, regardless of its membership in a particular speech community, differentiates between different types of events (and non-events). In other words, the human mind is capable of drawing a line between actions and non-actions, states and activities, events and processes. Disregarding the common principle of human cognition, languages, however, "essentially differ in what they must convey and not in what they may convey" (Jackobson 1959: 236). The fact that the Russian verb must convey the perfective-imperfective distinction implies that there is a distinction between the way languages manifest basic cognitive principles (universal), such as representation of time and space. For a Russian native speaker in particular, this representation is a fundamental dichotomy.

Recent accounts of Russian aspect, such as that of Kravchenko (2004), also propose to look at aspect from a cognitive perspective. Kravchenko (2004) argues, based on the morphological and syntactic evidence from Russian, that aspectual oppositions have little to do with "boundedness" and "totality" as it has often been claimed in formal descriptions of aspect (Smith 1991; Comrie 1976; Dahl 1985).

Instead, the choice between aspectual pairs is thus determined by the speaker's source of information about the event, i.e. the speaker's knowledge of the event and the speaker's observation of the event.

Below, we shall examine (i) markers, which are available for Slavic speakers for expressing temporal-aspectual relations, and (ii) functions these markers perform in the discourse of Slavic speakers of English. The use of tenses for non-past anchoring – the simple present, the present progressive and the present perfect – will be examined first. A consideration of tenses used for past-based anchoring, such as the simple past and the past progressive, will follow. The functions these tenses perform in the lingua franca context will be spelled out in the discussion.

2.1 The non-past-based anchoring

The simple present, the present progressive and the present perfect are tenses which are used to anchor the events in the non-past according to prescriptive grammar. In Standard English (Huddleston – Pullum 2002) the simple present is used to: 1) refer to events which habitually/regularly happen; 2) refer to events when the time is indeterminate and irrelevant as in giving facts; 3) refer to events which will happen in the future; 4) create an additional effect and bring the hearer/reader into the event, i.e. the historical present.

In the data, the simple present is the tense, which is widely used by all speakers in the study. Contrary to the functions of the simple present in Standard English, in the Eastern European variation of English, the simple present extends its functions and goes beyond the boundaries of simple present use. Below we shall consider some of the functions the simple present performs in the speaker data.

One of the functions of the simple present is to carry out a function of the present perfect. It is even used where the obligatory present perfect use is triggered by the temporal adverbials, such as *since*, *for a long time*, and *many times.* In (1), for example, the English temporal system requires placing the event *to be interested in politics* into the recent past. The Ukrainian speaker, however, does the opposite and only places the event in the present, given that it is allowed by the Ukrainian language.

(1) I started learning English when I was twelve at school, and I have a very strong motivation because, since my childhood I am interested in politics, in history. (L1 Ukrainian)

The simple present is also used by Slavic speakers to represent the duration of events, as for example, with the adverb *for a long time.* In Standard English, this function is normally performed by the present perfect. Consider this example from the performance data of a Ukrainian speaker of English:

(2) Of course, when I am doing my research, I am usually using my reading skills, reading skills, but I feel that I need some more oral practice, because when you live very long in your native country, so the language is forgotten. (L1 Ukrainian)

Another function of the English present perfect, namely, the use of the present perfect in the obligatory context with the temporal adverbial *now*, is

also performed by the present simple in the spoken data. In (3), the hearer is also faced with a tense shift, which makes it difficult to order the events.

(3) That was actually time when I had to catch my English, I have to improve, actually. I can't say that right now that I really improve it, but </break>. (L1 Ukrainian)

It is possible to account for the use of the simple present in the obligatory context of the present perfect by the difference the Slavic languages and English have in the threshold of the present time. The tense system of Russian, constructed with descending time, allows for the use of the non-past for events that began in the past and extend into the present (Hewson – Bubenik 1997: 333). The following sentence, for example, is possible in Russian:

Я уже говорю десять минут.
Ya uzhe govoru desyat minut.
I already speak ten minutes.
PRONOUN, 1st PS., SG., – VERB – PRESENT – ADV. – NUM. – NOUN – PL.

English, however, requires locating the event into the recent past[5]. Another extension of the simple present is rendering events in the past, i.e. in the obligatory context of the simple past. The excerpt in (4) illustrates the use of the simple present in the past-based anchoring by a Russian speaker of English:

(4) Yeah, but but today I can say that mhm I really find what I wanted. (L1 Russian)

Accounting for this type of occurrence of the simple present, it is possible to suggest that speakers have difficulties retrieving the past tense forms, which result, in turn, in the simplification of the temporal-aspectual system. Another tense, which locates the events in the present, additionally representing the internal consistency of the event, is the present progressive. In English, the category of the progressive is fully grammaticalized. In contrast to the events, which are used in the simple present and have a habitual reading, events in

5 This question has been examined in detail by Korrel (1991). She argues that the difference in the usage of the present perfect and the present simple in English and other I-E languages stems from a representation of the present as "just actualized" (Russian, German, Dutch) and 'not actualized' (English) as in **I speak for ten minutes*.

the present progressive have a single occasion interpretation and a dynamic character (Huddleston – Pullum 2002: 155). Dahl (1985) and Comrie (1976) claim that the English progressive is used in a wider range of contexts than progressives in other languages (Dahl 1985; Comrie 1976). It conveys more than a simple aspectual meaning. Bybee sees the function of the progressive in describing "subjects in the midst of doing something" (Bybee 1994: 35). Slavic languages, unlike English, do not have an obligatory progressive marking and the duration or progressivity of an action is inherently encoded into the verb semantics. Although, there is no obligatory marking for the progressive in Slavic languages, Slavic speakers tend to overuse this category.

If morphosemantic transfer cannot account for the emergence of this category, what is it that has an influence? A closer look at the data allows me to suggest the following: when marking verbs as progressive, the Slavic speakers do not intend to convey the predicate type/Aktionsart[6]; instead, they intend to convey the distinction between the Slavic perfective and imperfective, falsely associating the Slavic imperfective with the English progressive and the Slavic perfective with the English simple past. This assumption causes the use of the English progressive in the non-obligatory context, which leads to the over-extensive use of this aspectual marker.

Contrary to the temporal-aspectual constraints of English, the present progressive in the Eastern European manifestation is extensively used to convey habitual events and repetitive actions. Some occurrences of rendering habitual actions by the progressive aspect are presented below:

(5) I'm listening to the songs, and I'm reading, I try to read in English, and when it happens I try speak with people in English. (L1 Polish)

In (5) the Polish speaker intends to convey events, which happen on a daily basis, i.e. *the speaker reads books in English* and *listens to the music*. The use of the progressive aspect, therefore, is not obligatory here. It is also observed that the progressive aspect is used with the simple present triggers, such as temporal adverbials, which mark habituality and repetitiveness, for example, *usually*, *from time to time* as it is illustrated in the following example extracted from the performance of a Ukrainian speaker:

6 Vendler (1967) suggested a four-way categorization of verbs (states, activities, achievements and accomplishments) based on their semantics. This classification gave rise to further investigations of the effect the lexical aspect has in First and Second Language Acquisition. Bogdan – Sullivan (2009: 40) criticize Vendlerian categorization and propose a cognitive classification of Aktionsart that applies to both English and Polish.

(6) Of course, when I am doing my research, I am usually using my reading skills, reading skills, but I feel that I need some more oral practice, because when you live very long in your native country, so the language is forgotten. (L1 Ukrainian)

Here, the speaker uses the progressive aspect to speak about events which are habitual acts, and not actions in progress. Thus, the present progressive emerges in the obligatory context of the simple present, even when the use of the simple present is triggered by the temporal adverbial *usually*. Similarly, in the set of utterances which follow below (7-9), actions refer to habitual events and not to single occurrences that require the use of the simple present. Because the verbs *to speak* and *to try* are inherently durative in Russian, they emerge in the imperfective aspect. In (8) a Ukrainian speaker of English coins a verb *to maturitize* and uses it in the progressive aspect. Thus, speakers with L1 Russian and Ukrainian make use of the progressive aspect to convey the imperfective durative meaning as the examples (7-9) below illustrate:

(7) In Belarus, English is very useful. We are speaking Russian and second international language for us is English. (L1 Russian)

(8) You can learn eh some new words, but for people who are maturitizing as I think main thing is to have something interesting, some literature, or some text of the subject. (L1 Ukrainian)

(9) Well I am trying to participate in some international conferences, for example, in a few days, I will be in Istanbul at seminar. (L1 Ukrainian)

Thus, the present progressive is used with activities, as they are durative and unbound. Interestingly enough, not only activities are used in the progressive aspect. Even accomplishments which are bound (Vendler 1967), as it is illustrated in the set of examples (10-11) *to come from a small town* or as in *to come from Eastern Europe*, emerge in the progressive aspect. This, again, is not in accordance with the English temporal-aspectual system.

(10) I am coming from a small town, but I was I was studying in the village school, where my grandparents <break/> and <break/> living. (L1 Ukrainian)

(11) Everybody understood that we are coming from Eastern Europe, and actually Slovaks are also somehow at the same situation. (L1 Ukrainian)

As the progressive views action as ongoing at reference time it applies typically to dynamic predicates and not to stative ones (Comrie 1976). Slavic speakers of English, however, tend to extend the use of the progressive from dynamic predicates to stative ones (see above). Again, the underlying reason might be the speaker's willingness to render events which are imperfective in Russian. Stative verbs such as *to think* and *to feel,* for instance, do not tend to be used in the progressive aspect. In the performance of Slavic speakers, however, this is not the case. Consider the two sets of data excerpts with verb phrases *to think* in (12-13) and *to feel comfortable* in (14):

(12) They are thinking maybe I am Englishman you know. (L1 Russian)

(13) Maybe its sounds really not polite, but I am thinking that we better do some practical things then just to waste five years at university, without doing anything. (L1 Ukrainian)

(14) Its shows, that I no sometimes it depends, in some situation I am feeling comfortable and I have no problem with understanding, and in other situations, I have to ask more and more please, repeat. It's my worst feeling. (L1 Ukrainian)

Apart from the use of the progressive discussed above, it is used by Slavic speakers for rendering events which began in the past and were still going on at the time of utterance. The present progressive thus extends its functions to the domain of the present perfect progressive. Even when temporal adverbials such as *how long* and *over the last years*, which trigger the use of the present perfect progressive, are present, Slavic speaker use the progressive aspect only. The utterances in (15) and (16) that follow illustrate this:

(15) It's it's not easy to ask somebody oh how long are you studying English if she or he just started to study English, okay. (L1 Ukrainian)

(16) Mhm, yes, but it's a difference in how long they are studying language eh English language English language, and what is what is their using of English, eh mhm so. (L1 Ukrainian)

Now, let us examine the use of the present perfect in the speakers' production data. First, a closer look at the data shows that the present perfect is used by Slavic speakers when they recount events which happened at a definite point in time. In Standard English, however, this function is performed by the simple past. The excerpt in (17), for example, illustrates how a Ukrainian speaker narrates an event which commenced in the past and was over in the past, also specifying the time of the event. Thus, the event had no relevance for the utterance time.

(17) I've graduated this university in nineteen ninety nine, then I was a student of post-graduate program and I've, after that I defended my thesis. (L1 Ukrainian)

So far we have seen that non-native speakers of English tend to overextend the functions of the simple and progressive and use them in the obligatory contexts of other tense forms.

2.2 The past-based anchoring

The past-based tenses occurring in the data are the simple past and the past progressive. It was mentioned above that because aspect marking is obligatory in Slavic languages, Slavic speakers mark English verbs for aspect when constructing their past-based narrations. In doing so, Slavic speakers associate the Slavic unmarked imperfective with the English past progressive, and the Slavic perfective with the English simple past.

In what follows, I suggest taking a look at how these two past-based tenses are used by speakers in their narrations. Taking the data into account, it becomes possible to claim that in the Eastern European English variety, the simple past acquires the functions of the simple past and the present perfect. To illustrate this, let us consider the excerpt below:

(18) I hope I will get fractionation because one year ago, actually, not one year ago, this year, it was in March, when I participated in labs, and I did this actually liked this topic, because it's very interesting, and when we did it we didn't see any fractionation, we didn't do it properly, and my supervisor told me that it's also result, yeah, but, nevertheless, they will continue, and this probably we have to study more. (L1 Russian)

The speaker recalls the events in the past. She explicitly says that an action took place in the past, so the past tense seems to be acceptable and applicable for that. However, where there is a present time reference, the simple past is used to convey the function of the present perfect. In the following two excerpts (19) and (20), the events have a present time reference, which requires the use of the present perfect. Even the present perfect triggers, such as *now* and *for,* are ignored.

(19) Because now I finished institute, university, and I want to have some maybe Master. (L1 Ukrainian)

(20) It's normally, and I understood now because year by year, you just have to live to saw how life is, it's very important, and knowledges in books, they are also important, but knowing about life, just life. (L1 Ukrainian)

For rendering past actions which are imperfective in Slavic languages, the English past progressive is at the speakers' disposal. Similarly to the present progressive, it is used by Slavic speakers not to show the duration of an action but to convey the imperfective aspect in Slavic. Apart from instances where the past progressive is used in the past progressive non-obligatory context (in place of the simple past), it is also used in the context where the progressive aspect is unacceptable in English, as, for example, with predicate types which are neither durative nor unbound (what Vendler 1967 referred to as "achievements"). Examples in (21) and (22) extracted from the performance of a Ukrainian speaker illustrate this:

(21) I was travelling to Pakistan and <NLU> Arabsky Emiraty </NLU>, Dubai and even I have to speak, when I was speaking very good English language, or with good pronunciation, they can't understand me. (L1 Ukrainian)

(22) And sometimes, I was visiting some exhibitions in business and I try to speak English, because when I am speaking English they are more polite with me, they are more polite with me, they are thinking maybe I am Englishman you know. (L1 Ukrainian)

In the examples, the past progressive was used in the simple past obligatory context. Such predicate types as *to travel from Pakistan* as in (21) and *to*

visit exhibitions as in (22) are neither durative nor unbound in English. In Slavic languages, however, verbs *приезжать* (priezhat) 'to come', *путешествовать* (puteshestvovat) 'to travel', and *посещать* (poseshat) 'to visit' are imperfective. They, therefore, emerge in the progressive aspect in the speakers' performance.

3. Conclusion

This paper is restricted to a discussion of morphosemantic variables in the Eastern European English variety such as tense and aspect. Given that Slavic-like and non-Slavic-like temporal aspectual systems differ in how they construct the representation of events in time, the use of such morphosemantic categories as tense and aspect is likely to deviate from what one may find in Standard English. In general, a question was raised as to what it was that caused Slavic speakers to construct their temporal discourse in a certain way. The study drew on data from spontaneous spoken production of Ukrainian, Polish, Russian and Slovak speakers, compiled into a small-sized corpus of the Eastern European English variety.

Two general tendencies are observed here. First, as the category of aspect is a grammatical one, and the aspect marking is obligatory in Slavic, Slavic users of English tend to mark verbs as perfective and imperfective, falsely associating Slavic perfective with the English simple past, and the Slavic imperfective with the English progressive aspect. Second, as the English progressive is used by Slavic speakers for rendering imperfective actions, it emerges in the progressive non-obligatory context, which, in turn, leads to the overuse of the English progressive. The past progressive and the simple past are the basic tenses which render the past-based events. The non-past events are rendered by the simple present and the present progressive, where the present progressive conveys what is imperfective in Slavic.

In summary, Slavic speakers of English do not seem to fully use the available tense repertoire. Tense functions and boundaries become non-rigid and less fixed where the use is concerned; the English temporal-aspectual system in its Eastern European manifestation thus gets simplified and reduced. Based on what was mentioned above, it is possible to suggest that an Eastern European English(es) is indeed emerging. Additional studies, examining the domains of morphosemantics, morphosyntax and lexis, however, are necessary to provide further linguistic evidence.

REFERENCES

Sources

Tübingen Corpus of Eastern European English (TCEEE) (*University of Tübingen*).

Special studies

Albl-Mikasa, Michaela – Sabine Braun – Sylvia Kalina (eds.)
2010 *Dimensionen der Zweitsprachenforschung* [Dimensions of Second Language Research]. (A Festschrift for Kurt Kohn.) Tübingen: Narr Verlag.

Berns, Margie
1995 "English in Europe: Whose language, which culture?", *International Journal of Applied Linguistics* 5 (1), 21-32.
2005 "Expanding on the Expanding Circle: Where do WE go from here?", *World Englishes* 24 (1), 85-93.

Bogdan, David R. – William J. Sullivan
2009 *System of Polish Narrative. A Discourse and Cognitive Approach*. (LINCOM Studies in Slavic Linguistics.) München: Lincom.

Bolton, Kingsley
2003 *Chinese English: A Sociolinguistic History.* Cambridge: Cambridge University Press.

Bybee, Joan – Revere Perkins – William Pagliuca
1994 *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Comrie, Bernard
1976 *Aspect*. Cambridge: Cambridge University Press.

Comrie, Bernard – Greville G. Corbett
2002 *The Slavonic Languages*. London: Routledge.

Dahl, Östen
1985 *Tense and Aspect Systems*. Oxford: Basil Blackwell.

Durst-Andersen, Per
1994 "Russian aspect as different statement models". In: C. Bache et al. (eds.) *Tense-Aspect-Action*. Berlin: Mouton de Gruyter, 81-112.

Hewson, John – Vit Bubenik
1997 *Tense and Aspect in Indo-European Languages*. Amsterdam: John Benjamins.

Hopper, Paul J.
1977 "Aspect and foregrounding in discourse". In: T. Givon (ed.) *Syntax and Semantics 12: Discourse and Syntax*. New York: Academic Press, 213-241.

House, Juliane
2002 "Developing pragmatic competence in English as a lingua franca". In: K. Knapp, C. Meierkord (eds.) *Lingua Franca Communication.* Frankfurt: Peter Lang, 245-269.
Huddleston, Rodney – Geoffrey K. Pullum
2002 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press
Jackobson, Roman
1959 "On linguistic aspects of translation". In: R. A. Brower (ed.) *On Translation.* Cambridge, MA: Harvard University Press, 232-39.
Jenkins, Jennifer
2007 *English as a Lingua Franca. Attitude and Identity.* Oxford: Oxford University Press.
Kachru, Braj
1988 "The sacred cows of English", *English Today* 16, 3-8.
Kirkpatrick, Andy
2010 *The Routledge Handbook of World Englishes*. London: Routledge.
Knapp, K. – C. Meierkord (eds.)
2002 *Lingua Franca Communication*. Frankfurt: Peter Lang.
Korrel, Lia
1991 *Duration in English: A Basic Choice, Illustrated in Comparison with Dutch*. Berlin: Mouton de Gruyter.
Kravchenko, Alexander
2004 "A new cognitive framework for Russian aspect". In: F. Karlsson (ed.) *Proceedings of the 20th Scandinavian Conference of Linguistics*. Helsinki: University of Helsinki.
Lamb, Sydney M.
1991 "Linguistic Model and Linguistic Thought: the case of either-or thinking", *LACUS Forum* 17, 101-108.
Lim, Siew Siew – Ee L. Low
2005 "Triphthongs in Singapore English". In: D. Deterding et al. (eds.) *English in Singapore. Phonetic research on a Corpus*. Singapore: McGraw--Hill (Education) Asia, 64-73.
Meierkord, Christiane
1996 *Englisch als Medium der interkulturellen Kommunikation: Untersuchungen zum non-native-/non-native-speaker-Diskurs*. Frankfurt: Peter Lang.
Mendis, Dushyanthi – Harshana Rambukwella
2010 "Sri Lanka Englishes". In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London: Routledge, 181-196.
Mesthrie, Raj – Rakesh Bhatt
2010 *World Englishes. The Study of New Linguistic Varieties.* Cambridge: Cambridge University Press.
Moody, Andrew
2007 "English in Japanese popular culture and J-Pop music", *World Englishes* 25 (2), 209-222.

Mukherjee, Joybratto
2010 "The development of the English language in India". In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes.* London: Routledge, 167-180.
Proshina, Zoya
2010 "Slavic English: Education or culture?" In: A. Kirkpatrick (ed.) *The Routledge Handbook of World Englishes*. London: Routledge, 299-316.
Proshina, Zoya – Bryan Ettkin
2005 "English-Russian language contacts", *World Englishes* 24 (4), 439-444.
Salakhyan, Elena
2012 "The emergence of Eastern European English", *World Englishes* 31 (2), 270-289.
Schneider, Edgar
2003 "The dynamics of New Englishes: From identity construction to dialect birth", *Language* 79 (2), 233-281.
Seidlhofer, Barbara
2010 "Lingua Franca English. The European context". In: A. Kirkpatrick (ed.)
The Routledge Handbook of World Englishes. London: Routledge, 355-371.
2004 "Research perspectives on teaching English as a lingua franca", *Annual Review of Applied Linguistics* 23, 209-39.
Seidlhofer, Barbara – Henry Widdowson
2009 "Conformity and creativity in ELF and learner Englishes". In: M. Albl-Mikasa et al. (eds.) *Dimensionen der Zweitsprachenforschung. Dimensions of Second Language Research.* (Festschrift for Kurt Kohn.) Tübingen: Narr Verlag, 97-103.
Simo Bobda, Augustin
2000 "The uniqueness of Ghanaian English pronunciation in West Africa", *Studies in the Linguistic Sciences* 30 (2), 185-98.
Smith, Carlota S.
1991 *The Parameter of Aspect*. Dordrecht: Kluwer.
Takeshita, Yuko
2010 "East Asian Englishes. Japan and Korea". In: A. Kirkpatrick (eds.) *The Routledge Handbook of World Englishes*. London: Routledge, 265- 281.
Ustinova, Irina P.
2006 "English and emerging advertising in Russia", *World Englishes* 25 (2), 267-27.
Vendler, Zeno
1967 *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

On the complementation of *start, begin* and *continue* in spoken academic American English

Oskar Gawlik
Jan Kochanowski University

ABSTRACT

The main goal of this article is to present the results of a linguistic inquiry into the syntax of *start*, *begin* and *continue* in contemporary spoken academic American English. The main goal of the study was to determine the frequency of occurrence of these verbs and their preferred non-finite complements – the gerund form and the present infinitive form – in that variety of English. The incidence of *start, begin* and *continue* was investigated in both small and large lectures in the MICASE corpus. The precepts of corpus linguistics have together served as the major methodological tool. The audience targeted here includes theoretical and applied linguists interested in English linguistics as well as students of the discipline and scholars of related fields.

1. Introduction

There are several good reasons for investigating the preferred complements of *start*, *begin* and *continue* (i.e. the gerund and the present infinitive) in contemporary academic spoken American English. Generally, variation is natural to language, and genre-induced variation in English deserves more attention than it has so far received. More specifically, genre-induced variation in the syntax of the lemmas of *start, begin* and *continue* in spoken academic English, due to genre-specific lexicogrammatical patternings, is one particularly interesting yet apparently neglected area. Most English grammar handbooks lack substantive information concerning the preferred complements of *start, begin* or *continue.* Moreover, such references say little or nothing on genre-specific preferences for the non-finite complements of these verbs (i.e. the gerund and the full infinitive), the exception being

the *Longman Grammar of Spoken and Written English* (Biber et al. 1999). In the absence of descriptive detail, we might assume that these verbal complements are in free distribution irrespective of genre, a notion which has for present purposes been assumed not to be the case. Consequently, striving for naturalness of expression, for example, when delivering an academic paper, and unable to answer my own or my students' questions concerning the most frequent, ergo most natural, verbal complements of *start, begin* and *continue* in spoken academic English, I embarked on a study, a corpus study, into the syntax of these verbs, an enterprise which I, and others, such as Swales (1990), the doyen of English for Academic Purposes (EAP), have resorted to since English grammars and other sources fail to provide unequivocal answers. Furthermore, an attempt has been made to investigate spoken academic English, instead of written academic English, on the grounds that the latter has been explored relatively thoroughly, while the former still appears to be *terra incognita*.

2. Corpus linguistics as a methodological tool

Although corpora were drawn on to investigate language even in the heyday of Chomsky's generative-transformational grammar, corpus linguistics (CL) offers a relatively new methodological tool. It enables us large scale investigation of actual linguistic production, rather than investigation of the results of linguistic introspection, which has not infrequently proven unreliable. CL appears to represent a methodological shift from a focus on competence to one on performance, to use Chomsky's terms, or from one on *langue* to one on *parole*, to employ de Saussure's terms. However, the swing of the pendulum has not applied across the board: a number of scholars advocate complementing corpus-based and corpus-driven methodologies with introspection, particularly in qualitative analyses (e.g. Rusiecki 2006), an approach I also adopt in my linguistic inquiries.

The value of corpus linguistics and its empirical basis is difficult to overestimate, and the wide panoply of linguistic pursuits that it involves results in numerous implications and applications of findings, particularly because of the burgeoning number and variety of corpora being made available. However, in the interests of space, I will not attempt a comprehensive treatment of the merits of CL here, but rather will simply laud one major work which represents the methods and fruits of CL: the *Longman Grammar of Spoken and Written English* (LGSWE; Biber et al. 1999).

This grammar is devoted to an analysis of lexicogrammatical patternings in four genres of current English: academic, journalistic, fictional, and conversational. Its quantitative and qualitative analysis of the characteristics of these four English genres paints a comprehensive picture of English grammar, highlighting the differences between the genres and validating previous (hypo)theses concerning language while disconfirming others, which, understandably, has a number of implications and applications, pedagogical ones among them. Consequently, LGSWE appears to have surpassed other books devoted to the grammar of English inasmuch as it both epitomizes the grammatical theory of CL and (re)validates CL's empirical approach, this last by reference to actual language use, which reveals the employment of certain linguistic structures while eschewing others, genre by genre. No doubt interest in employing CL as a methodological tool is gaining popularity. Tens of English language corpora are being utilized by scholars worldwide today, and increasing numbers of corpus-based or corpus-driven publications are appearing (cf. e.g. Mauranen 2001; Swales 2001; Łyda 2007; Gawlik 2011).

3. Selected studies of academic English

As concerns academic English, a large number of studies have been done in applied linguistics, with practical applications constituting the main rationale for discovering the intricacies of academic discourse. Consequently, linguistic interests in academic English have mostly revolved around the study of written English grammar and lexis, such as those of 'the research article', the academic written genre *par excellence*, with a view to exploring a wide range of linguistic aspects: contrastive rhetoric (Galtung 1981; Clyne 1987; Scollon – Scollon 1995; Connor 1996; Mauranen 1997), genre analysis (Swales 1990), metadiscourse (Hyland – Tse 2004), hedging (Hyland 1998; Varttala 2003) or evaluation (Stotesbury 2003). The results of such linguistic inquiries are often extended to practical applications: 'the research article' is a genre which millions of people, both students and academics, are exposed to and grapple with in the milieu of academia. In sum, explorations into academic written discourse, frequently propelled by a view to practical applications, appear to have borne fruit in the form of numerous works targeted at both experts and student neophytes.

Of late, however, investigations of academic English have bifurcated into analyses not only of the written mode of academic discourse but

also those of the equivalent spoken mode. The latter have stemmed from both a growing cognizance of the need to expand such studies and an increased availability of corpora of spoken academic English, such as the *Michigan Corpus of Academic Spoken English* (MICASE; Simpson et al. 2002). Consequently, research into EAP, particularly into the prototypical genre of spoken academic English, or 'the lecture talk', as exemplified by studies of metatext (Swales 2001), reflexive academic talk (Mauranen 2001), evaluation (Swales 2004), concession (Łyda 2007), and *verba dicendi* (Gawlik 2011), has gained momentum. Still, despite a growing body of research, investigation of the spoken variety of academic English remains in its infancy, which speaks to importance of the present study.

4. Methods of analysis of the occurrence of the non-finite verbal complements of *start*, *begin* and *continue*

The focus in this investigation was on the nearly synonymous verbs *start* and *begin* as well as the verb *continue*, and the patterns of verbal complementation regarding the apparently competing gerundial and infinitival non-finite complements, with a view to determining which of the two is more preferable in spoken academic English. The choice of these two types of complements precluded analysis of any other types of complements, such as the progressive or perfect infinitives. That decision was made for at least two reasons: firstly, the gerund and the infinitive mark a different aspect, and thus convey different nuances of meaning, which implies that they should not be considered complements that are competing with each other in relation to the three aspectualizers analysed. The disregard for other types of complementation also stemmed from the supposition that the results of the study could have specific practical, perhaps pedagogical, applications concerning relevant verbal complement choices.

The synchronic corpus investigation was conducted on the prototypical genre of spoken academic English, the lecture talk, and relied on data from MICASE, a collection of authentic and unscripted texts which were recorded at the University of Michigan and which consequently represent the American variety of English. The corpus consists of almost 200 hours (approximately 1.8 million words) of contemporary academic speech, divided into sixteen different speech events, encompassing topic-matter of both the humanities and the sciences. The corpus can be accessed using a wide variety of sociolinguistic filters. For instance, one may search

for data in speech events of women only or men only. Unfortunately, for all the merits of the corpus, it is not tagged for grammatical categories, which meant that the data culled needed to be checked manually for categorical sorting in order to avoid arriving at skewed results.

The MICASE data was complemented by data of the *Corpus of Contemporary American English* (henceforth COCA) complied by Mark Davies at Brigham Young University in 2008. That corpus is the largest contemporary body of American English, and it spans the years 1990–2011. COCA is annotated biannually, and currently it consists of approximately 425 million words of five genres: spoken, fictional, journalistic-magazine, journalistic-newspaper, academic. Grammatically tagged, dialectally representative, inclusive of spoken academic data, comparatively large, the corpus was invaluable to the present study.

5. Previous inquiry into complementation

By now, a number of studies on complementation, both synchronic and diachronic, have employed the methodology of CL (cf. e.g. Rudanko 2000; Egan 2008; Mair 2009). However, before CL began to develop in earnest, statements about complementation which were based on introspection and few instantiations were, to an extent understandably, the norm. One such statement regarding two of the aspectualizers examined, *start* and *begin*, is this:

> An informal survey of native speakers of English indicates that most believe *begin* and *start* to be close synonyms and almost entirely interchangeable. Some say only that *begin* seems slightly more 'formal'. This feeling may be due to the fact […] that *begin* occurs in a more restricted number of contexts than *start*. A careful analysis of these two aspectualizers, however, turns up a surprising number of differences […]. [T]here must exist semantic as well as syntactic distinctions between these aspectualizers which native speakers attest to by their unselfconscious and natural use of them [sic.]" (Freed 1979: 68).

The claim here that the two verbs are virtually interchangeability is typical of traditional grammar handbooks and usage guides. More interesting is the hypothesis that the number of contexts of the respective nouns differs, and the suspicion that semantic and syntactic restrictions must also affect the relevant

behavior of these verbs (and allow them to avoid obsolescence). It is from the point of such suppositions that we may proceed by the means of CL and attempt to establish patterns of restriction and preference, patterns among instances of non-finite verbal complementation specific to these verbs (and, for present purposes, *continue* as well), with a greater measure of certainty.

6. The syntax of *start*, *begin* and *continue* in light of empirical investigations

The synchronic corpus investigation of the gerundial and infinitival complementation of the lemmas of *start*, *begin* and *continue* in the MICASE, yielded the results included in Appendix 1. The pertinent data there, after lemmatization, include all and only verbal occurrences of the word forms, and all and only constructions containing non-finite gerundial or infinitival complements. Because the MICASE set of lectures, small and large, consists of 584,970 words, the figures in Appendix 1 are normalized to one million occurrences for ease of interpretation. For ease of comparison, the results are also presented in Appendix 2 in a graphic form.

Statistically, *start* is the most frequent of the verbs, and it is most often followed by a gerundial complement, a total of 171 times (approximately 292 occurrences per million words). By comparison, infinitival complements of *start* are less numerous, totaling 120 instances (approximately 205 occurrences per million words). Items (1) and (2) below exemplify the two respective types.

(1) you had to memorize? okay, you have a lot of A-T-P molecules when you're living on glucose. if you **start** depleting the glucose you're gonna be running out of A-T-P and available A-T-P will get converted in. (LES175SU079)

(2) assigned to that descriptor. um, i think, here experience comes to bear. i think after a while you **start** to see certain descriptors, become familiar with them, see them more and more, and find, ah this will. (LES335JG065)

Begin, by contrast, exhibits a weak preference for the gerund, with only 7 instances (12 occurrences per million words), and a strong preference for the infinitive, with as many as 84 instances (143 occurrences per million words). Examples of those instances are seen here:

(3) thousands can vote, and people are rigging votes in all kinds of exciting and interesting ways, this **begins** to matter. there is a feeling that, you never get to vote. your vote never counts. what ma- who care. (LEL215SU150)

(4) and federal governments, personnel officers, they **begin** entering graduate programs. they'd already **begun** entering medical schools already eighteen fifties uh som- women and law schools becomes i. (LEL105SU113)

Continue, like *begin*, is more frequently complemented by the present full infinitive than it is by the gerund. The figures, respectively, are these: 46 occurrences (79 per million words) versus 4 occurrences (7 per million words). Examples of each follow.

(5) than a hundred miles of country before reaching the sea, into which they plunge, unhesitatingly, and **continue** to swim on until they die. even then they float so that their dead bodies form drifts, on the seasho. (LEL175SU112)

(6) right but why can't you just say sit still, and, **continue** making money (xx) cuz somebody. (LEL 565SU064)

As these results indicate, when set beside *begin, start* is the more frequent word (near-synonym) in spoken academic American English. Not distantly related to the respective frequencies of these two verbs is a pattern regarding comparative formality. The fact that both words occur in this corpus of formal English, and manifestly in not wildly different quantities of instances, argues against a significant difference in usage in terms of formality, and against the supposition by Freed (1979: 68) quoted above. Of course the time of a generation has passed since that supposition was stated. The results also suggest that *start* is typically complemented by the gerund form, while *begin* is typically complemented by the full infinitive, with the gerund form appearing to constitute an exception.

My corpus findings tally with those of other reseachers with respect to *begin* as "[one of] the most common verbs controlling to-clauses", and *start* as "[one of] the most common verbs controlling *ing*-clauses" (Biber et al. 1999: 699). However, as regards the complements of *continue*, my findings stand in contrast. Biber et al. (1999: 741) claim that "[a]apectual verbs are common

with *ing*-clauses, because their meanings concern the delimitation of actions that go on over time – relating to the starting point (e.g. *start*, *postpone*), the end point (e.g. *stop*, *quit*), or their progress (e.g. *keep on*, *continue*, *resume*)". This general claim does not appear to be applicable to spoken academic American English, as my corpus findings testify to the contrary: *continue* is typically complemented by the present to-infinitive in the MICASE lectures, which, by implication, might constitute a style marker of spoken academic English.

Although my original intention was to investigate four, rather than three, aspectual verbs, the fourth being the semantically proximal *commence*, on searching for tokens of *commence* in the MICASE, it emerged that there was not a single occurrence of any form of this verb either in the lecture talk analysed or in any other speech event contained in the corpus, which, given the verb's comparatively formal connotation, was unexpected. Since this fact ran counter to my introspective judgement, I sought evidence of *commence* and its complementation in a much larger corpus of American English, i.e. the *Corpus of Contemporary American English* (COCA), which currently consists of approximately 425 million words, in order to amass some genre-varying data on the verb and to confirm or repudiate the relatively low frequency of the verb evident in the formal MICASE data. Appendix 3 illustrated the COCA data culled.

There are at least two obvious conclusions to be drawn from the results included in Appendix 3. First, *commence* is of comparatively low general frequency, ranging from 1.3 to 6.96 tokens per one million words. Second, *commence* was more frequently employed in written academic English than it was in any of the other four genres, scoring 597 hits here, which translates into its 6.96 tokens per one million words. Of course, this implies little or nothing about its relative popularity in spoken academic American English. However, the tendency might suggest an overcoming of field and tenor, to frame the statement in Hallidayan parlance. Although the results need validation, what the corpus findings appear to have corroborated is that *commence* is infrequent not only in spoken academic English (as attested by MICASE), but also in certain other genres, as confirmed by the findings obtained from COCA. Although we might reasonably expect to encounter the comparatively formal verb *commence* frequently in formal spoken academic English, followed there by *begin* and *start*, in descending order of incidence, the evidence gleaned indicates that spoken academic American English discourse prefers *start* before *begin*, and both of these aspectualizers before

that of *commence*, and by extension, then, suggests that spoken academic discourse lies, in terms of formality, between written academic discourse and everyday spoken discourse, a pattern identified elsewhere (Swales 2004).

7. Conclusions

In light of the evidence presented above, it may be concluded on grounds of their respective frequencies of occurrence and patterns of complementation that the two aspectualizers *start* and *begin* are not absolute, but rather proximate, synonyms of spoken academic American English. In terms of frequency, *start*, with 291 hits in MICASE, is clearly a more popular option than *begin*, with its 91 hits. As regards syntax, *start* selects gerundial complements (171 hits in MICASE) more commonly than it does infinitival ones (120 hits), and *begin* manifests the opposite pattern, selecting the full infinitive (84 hits) far more often than the gerund (7 hits). The verb *continue* also prefers the infinitival complement (46 hits) to the gerundial (4 hits) in MICASE (see Fig. 1).

These differences in frequency and syntax may imply contrasting style marking by *start* and *begin* in spoken academic English. Investigation of the overall incidence of these two verbs, conducted from semantic and pragmatic viewpoints, would certainly further our understanding of their behavior in more formal as well as less formal contexts.

REFERENCES

Sources

Corpus of Contemporary American English (COCA) Compiled by Mark Davis (Brigham Young University). http://corpus.byu.edu/coca/.

Simpson, R.C. et al.
2002 *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Special studies

Biber, Douglas et al.
1999 *Longman Grammar of Spoken and Written English*. Essex: Longman.

Clyne, Michael
1987 "Cultural differences in the organization of academic texts", *Journal of Pragmatics* 11, 211-247.
Connor, Ulla
1996 *Contrastive Rhetoric. Cross-Cultural Aspects of Second-Language Writing.* Cambridge: Cambridge University Press.
Egan, Thomas
2008 *Non-Finite Complementation: A Usage-Based Study of Infinitive and -ing Clauses in English*. Amsterdam: Editions Rodopi.
Freed, Alice F.
1979 *The Semantics of English Aspectual Complementation.* Dordrecht: Reidel Publishing Company.
Galtung, Johan
1981 *Structure, Culture and Intellectual Style: An Essay Comparing Saxonic, Teutonic, Gallic and Nipponic Approaches*. London: SAGE.
Gawlik, Oskar
2011 *Basic Verba Dicendi in Academic Spoken English. The Syntax, Semantics,* Pragmatics and Functions of Communication Verbs: Say, Talk, Tell and *Speak.* Saarbrucken: Lambert Academic Publishing.
Hyland, Ken
1998 *Hedging in Scientific Research Articles.* Amsterdam: John Benjamins.
Hyland, Ken – Po Ting Tse
2004 "Metadiscourse in academic writing: A reappraisal", *Applied Linguistics* 25(2), 156-177.
Łyda, Andrzej
2007 *Concessive Relation in Spoken Discourse*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
Mair, Christian
2009 *Infinitival Complement Clauses in English. A Study of Syntax in Discourse.* Cambridge: Cambridge University Press.
Mauranen, Anna
1997 *Cultural Differences in Academic Rhetoric.* Frankfurt: Peter Lang.
2001 "Reflexive academic talk: Observations from MICASE". In R.C. Simpson et al. (eds.) *Corpus Linguistics in North America. Selections from the 1999 Symposium.* Ann Arbor: University of Michigan Press, 165-178.
Rudanko, Martti Juhani
2000 *Corpora and Complementation: Tracing Sentential Complementation Patterns of Nouns, Adjectives and Verbs over the Last Three Centuries.* Lanham, Maryland: University Press of America.
Rusiecki, Jan
2006 "Badania korpusowe a 'lingwistyka gabinetowa': czy te dwa podejścia kiedyśsię spotkają?". In A. Duszak et al. (eds.). *Korpusy*

w angielsko-polskim językoznawstwie kontrastywnym. Kraków: UNIVERSITAS, 17-29.

Scollon, Ronald – Suzanne Scollon

1995 *Intercultural Communication: A Discourse Approach.* Oxford: Blackwell.

Stotesbury, Hilkka

2003 "Evaluation in research articles abstracts in the narrative and hard sciences", *Journal of English for Academic Purposes* 2, 327-341.

Swales, John

1990 *Genre Analysis. English in Academic and Research Settings.* Cambridge: Cambridge University Press.

2001 "Metatalk in American academic talk. The cases of *point* and *thing*", *Journal of English Linguistics* 29(1), 34-54.

2004 "Evaluations in academic speech: First forays." In G. d. E. Camiciotti and E. T. Bonelli (eds.) *Academic Discourse – New Insights into Evaluation.* Bern: Peter Lang, 31-53.

Varttala, Teppo

2003 *Hedging in Scientifically Oriented Discourse. Exploring Variation According to Discipline and Intended Audience*. Ph.D. dissertation. Tampere: University of Tampere, Finland.

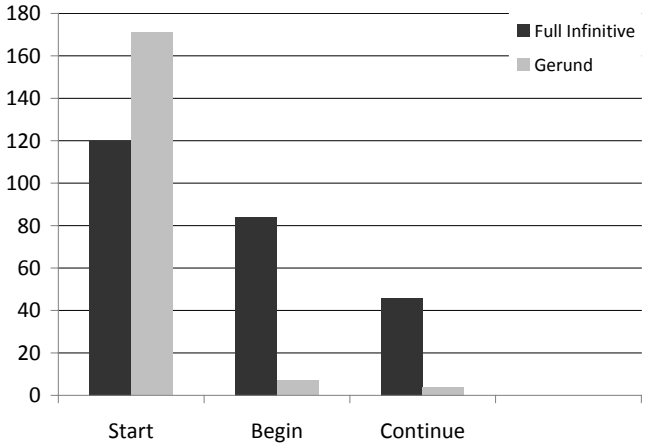
APPENDIX 1

The gerund versus the infinitive as complements of *start*, *begin* and *continue* in the MICASE corpus of lectures

The verb investigated	Total occurrences	Occurrences after lemmatization	Complementation by means of the full infinitive		Complementation by means of the gerund	
			Absolute number of tokens	Number of tokens per one million words	Absolute number of tokens	Number of tokens per one million words
Start	654	291	120	205	171	292
Begin	262	91	84	143	7	12
Continue	104	50	46	79	4	7

APPENDIX 2

The occurrences of the gerund versus the full infinitive as complements of the tokens of *start*, *begin* and *continue* in the MICASE corpus of lectures.



APPENDIX 3

Deployment of tokens of *commence* in different genres in contemporary American English (COCA data).

Genre	Total number of tokens	Frequency of occurrence per one million words
Spoken	117	1.3
Fiction	497	5.85
Magazine	335	3.71
Newspaper	187	2.16
Academic	597	6.96

Size-sound symbolism in the names of cars

Łukasz Stolarski
Jan Kochanowski University

ABSTRACT

This paper discusses the theory of size-sound symbolism which predicts that certain articulatory and/or acoustic characteristics of speech sounds have the potential to symbolise varying degrees of size of different objects. In particular, it examines the extent to which the assumptions of the theory have been applied in the process of creating car names. A set of 260 names of models of cars produced by various manufactures has been analysed phonetically and the results obtained have been juxtaposed with the volume of the respective vehicles. The conclusions reached in the study confirm that size-sound symbolism is utilised in brand names, but only to a limited extent.

1. Introduction

Sound symbolism, also referred to as "phonetic symbolism" (Sapir 1929, Newman 1933, Brown et al. 1955, Marchand 1960) or "phonosymbolism" (Malkiel 1994), is defined as "a general term for an iconic or indexical relationship between sound and meaning, and also between sound and sound" (Åsa 1999: 4). Even though the discussion of this topic can be traced as far back as Plato (cf. the discussion in Åsa 1999, Klink 2000, Yorkston and Menon 2004, Lowrey and Shrum 2007), the issue has been of particular research interest since the 20th century. To provide a few examples of works, some authors have conducted experiments on natural words (cf. Brown et al. 1955, Maltzman et al. 1956, Brackbill and Little 1957, Wichmann et al. 2010, Urban 2011) or nonce words (cf. Sapir 1929, Newman 1933) and analysed their phonetic structure in search of any associations with the words' meanings. There is also growing literature on the notion of "phonestheme" (cf. Bolinger 1950, 1965, Markell and Hamp 1960, Marchand 1960, Jakobson

and Waugh 1979, McCune 1985, Nordberg 1986, Rhodes 1994, Åsa 1999) which is understood to be a cluster of phonemes (or, sometimes, a single segment) appearing in words that are semantically, though often not etymologically, related. Furthermore, many authors have tried to establish a possible connection between meaning and individual phonemes. One can find, for instance, numerous studies on poetic language, in which various texts are analysed in terms of the sound-symbolic potential of consonants and vowels (cf. Tolman 1906, Lucas 1955, Householder 1960, Hymes 1960, Murdy 1966, Nash 1980, Chapman 1982, Frazer 1982, Caltvedt 1999, etc.).

As summarised in Stolarski (forthcoming), various typologies of sound-symbolism have been proposed. For instance, Åsa (1999) suggests a division into onomatopoeia, expressive interjections and sound-symbolic phonesthemes, while Marchand (1960) mentions direct imitation and expressive symbolism. One of the most elaborate classifications has been put forward in the preface to a collection of papers edited by Hinton et al. (1994). The notion is divided into corporeal sound symbolism, referring to a special use of sounds and intonation to express inner physical and emotional states, imitative sound symbolism, related to "onomatopoeic words and phrases representing environmental sound" (Hinton et al. 1994: 3), synaesthetic sound symbolism, focusing on "the acoustic symbolisation of non-acoustic phenomena" (Hinton et al. 1994: 4) and conventional sound symbolism dealing mostly with phonesthemes.

Of particular interest to this study is synaesthetic sound symbolism, especially its subtype known as "size-sound symbolism", or "magnitude sound symbolism" (Nuckolls 1999), which aims to look for associations between certain acoustic and/or articulatory characteristics of speech sounds and the semantic field of "size". One of the first authors to tackle this issue was Jespersen (1922) who suggested that in many languages the vowel /i/ is frequently associated with entities which are small, weak or insignificant. The idea was supported by Sapir (1929) who demonstrated that there is a correlation between the degree of openness in the articulation of vowels and the semantic dichotomy "big-small". This theory was further developed by authors such as Newman (1933), Bentley and Varon (1933) and Nichols (1971), and it has become commonly accepted that in the majority of languages open and/or back vowels are perceived as "bigger" than close and/or front vowels.

A possible explanation for such a perception of vocalic articulation may be associated with, as Sapir (1929) calls it, the "kinesthetic" factor. While pronouncing high vowels the tongue is raised towards the palate and the

resulting resonance chamber is small. The speaker subconsciously associates this close approximation with the notion of "small size". In the case of open vowels the situation is reversed and this is why they are more "appropriate" to symbolise objects of "large size". An alternative explanation proposed by Sapir involves the claim that "the inherent 'volume' of certain vowels is greater than that of others" (1929: 235). This intuitive idea was developed by Ohala (1983, 1984, 1994) who calls his theory "frequency code". The basic claim he makes is that high acoustic frequency is associated with the meaning "small vocalizer" and low acoustic frequency is connected with the primary meaning "large vocalizer". He states that: "In consonants, voiceless obstruents have higher frequency than voiced because of the higher velocity of the airflow, ejectives higher than plain stops (for the same reason) and dental, alveolar, palatal and front velars higher than labials and back velars. In the case of vowels, high front vowels have higher F_2 and low back vowels the lowest F_2" (1984: 9).

It is worth adding that the basic assumptions concerning the size-sound symbolic potential of vowels were confirmed in a cross-linguistic study by Ultan (1978). As mentioned in Stolarski (forthcoming), Ultan analysed as many as 136 languages and the results of his research confirm that diminutive size is frequently associated with high and/or front vowels.

In recent years the theory of sound symbolism has been applied in several studies on brand name development. Of particular interest to the present paper is the study by Klink (2000) who investigated, among other things, the perception of front versus back vowels in potential new brand names. The results he obtained fully confirmed the assumed association of such articulations with diminutive size. Moreover, Lowrey and Shrum (2007) conducted a psycholinguistic experiment in which they showed that in potential brand names for two-seater convertibles words with front vowels were preferred over words with back vowels. Conversely, in potential names for sport utility vehicles words with back vowels were found more fitting than words with front vowels. These results closely correspond to the assumptions of size-sound symbolism because two-seater convertibles are substantially smaller than SUVs. Additionally, Lowrey and Shrum (2007) frequently refer to the idea advanced by Yorkston and Menon that "consumers use information they gather from phonemes in brand names to infer product attributes and to evaluate brands" (2004: 43). Yorkston and Menon also suggest that the process in which sound symbolism manifests in consumer judgements is "uncontrollable, outside of awareness, and effortless, making it automatic" (2004: 43).

The major aim of this study is to examine the degree to which size-sound symbolism is applied in the creation (and possibly the consumer interpretation) of real brand names. In particular, the distribution of vowels will be investigated. The sample chosen in the experiment described below involves names of cars because they can easily be juxtaposed with the size of the vehicles they refer to. Additionally, even though many cars are not produced in an English speaking country, their names are usually either natural English words or are designed in such a way as to be pronounceable in English for reasons of broad marketability. It must be stressed, though, that the study does not aim at evaluating the assumptions of the theory itself. It is not designed to estimate the possible size-sound symbolic potential of English phonemes and to verify the claims discussed above. The aims are limited only to assessing whether or not the theory is applied in the process of car name-giving.

2. Data and methods

In order to investigate the problem outlined in the introduction, data on the sizes of 260 cars have been collected. In this process the following limits were adhered to:

1. The vehicles taken into consideration were models produced between 2006 and 2011. In this way the issue of technological process was avoided. Technical specifications of cars, including their size, have undergone constant change since the beginnings of the automotive industry. It is likely that the idea of a big car may be different now than it was, for example, 30 years ago. However, such a factor should not have any considerable influence within the period of 5 years chosen for the present study.

2. Only those cars whose names could be pronounced as "normal" English words were included. Abbreviations and acronyms were automatically excluded. Even if their articulation involves vowel sounds, they focus buyers' attention mainly on the consonants. Consequently, they are not suitable for examining the size-sound symbolic potential of vowels.

3. In most cases which involved two-word names only the first was analysed phonetically. The second word was frequently used to distinguish between several models produced by a given manufacturer

and expressed additional qualities common to different cars. For example, in the case of Toyota "Fortuner 3.0D-4D Automatic 2010" only the word "fortuner" was examined, since the element "automatic" is commonly used with other models, for instance in "Corolla 1.8 Automatic 2011", "FJ Cruiser 4×4 Automatic 2011", "Matrix S Automatic 2011", "Yaris 1.5 Automatic 2011", and so forth.

4. A given model is almost always available in more than one version. This immediately caused problems with establishing the volume of vehicles, because it was frequently the case that the size of one version differed sharply from the size of another. In order to tackle this issue the following solution was applied: in the case of each model the smallest possible version was chosen. This allowed a relatively objective comparison of different models. If the smallest version of one car was bigger than the smallest version of another car, then the former model could be regarded as generally bigger than the latter.

5. In each case, the size of the vehicles was established as a product of three values: the length, the width and the height. Obviously, such a solution fails to recognise differences in car shapes and the value ascribed to each vehicle is only an approximation of its overall volume.

The data gathered for the current experiment are presented in the appendix. They are based on the information found on various Web sites. The two which were most frequently accessed were "carsplusplus.com" and "autos.yahoo.com", but occasionally other web pages were also searched. It must be stressed that the vehicles listed in the appendix are not all the possible models for the period 2006 to 2011. Indeed, the total number of cars produced during this time is possibly substantially higher. Nevertheless, considerable effort has been invested in collecting data on models produced by diverse car manufacturers. In the appendix one can notice cars from such countries as the USA, Japan, England, France, South Korea, Italy, Germany, Spain and Romania.

As far as the RP pronunciation of the names is concerned, it was, first and foremost, established on the basis of Wells (2008). In cases when a given name was not found in Wells, it was deduced from the pronunciation of similar words commonly used in English. For instance, the articulation of "Fluence" was established on the basis of its similarity to the word "fluent", and the possible pronunciation of "Espace" was formed by comparing it to "escape"

and "space". In some more complicated cases, the solutions were found on various internet forums where the problem of the "correct" articulation of a given name was discussed. It must be stressed, however, that some names involve a degree of variability in the way they are pronounced. There may be consistent differences in the manner they are articulated in different dialects of English. Such discrepancies result from systemic, realisational, lexical or distributional differences (cf. Gimson 1994: 81-82). For example, Chevrolet Aveo is relatively consistently articulated as /ə'veɪoʊ/ in General American, but as /'æviəʊ/ in RP. In the present study this problem has been solved by choosing only one dialect for the analysis. Nevertheless, even in RP itself, some of the car names in the appendix may involve alternative articulations. Therefore, the suggested transcription should be treated as an attempt to provide the most typical pronunciations of the names, although in some cases other phonetic realisations are also possible.

As noted in the introduction, there are reasons to assume that back and/or open vowels are more suitable to represent small objects, while front and/or close vowels are expected to be associated with the notion of "large size". So far no convincing evidence has been found as to which of the two articulatory scales – the vertical or the horizontal – is more crucial in size-sound symbolism. Therefore, in the experiments described below both scales have been treated as equally important. With this in mind, the RP vowels have been divided into the ones which theoretically have the potential to symbolise "small size" and those which are more suitable in names referring to "large size". The result of this initial analysis, based on the RP vowel qualities suggested in Gimson (1994), Wells (2008) and Roach (2009), are depicted in Figures 1 and 2.

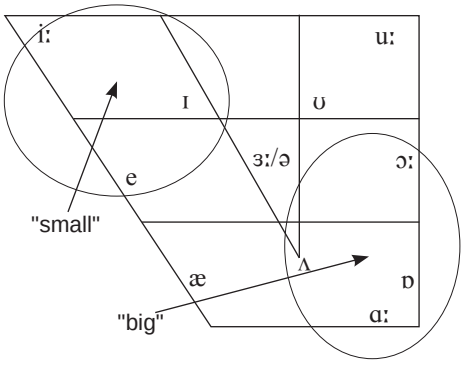


Figure 1. RP monophthongs and their predicted relation to the notions of "small" and "big"

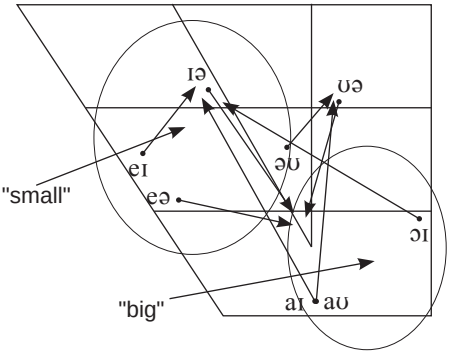


Figure 2. RP diphthongs and their predicted relation to the notions of "small" and "big"

Out of the 12 RP monophthongs the ones which are presumed to be "small" are /i:/ and /ɪ/ (cf. Figure 1). They are both high and front, which suggests their strong size-sound symbolic potential. The case of the vowel /e/ is less obvious. Although it is clearly front, it is also pronounced in the middle of the vertical articulatory scale. The potential of the vowel to symbolise "small size" is, consequently, less certain than with /i:/ and /ɪ/.

In many types of phonemic transcription, including the one used here, one may also encounter [i], which is intended to represent the final vowel in words such as *pretty*, but without having its own separate phonemic status. Gimson refers to such an articulation as "a short variety of /i:/" (1994: 99) and explains that no ambiguity arises from such pronunciations since the contrast between /ɪ/ and /i:/ is neutralised in word-final position. It should be stressed that in both experiments described below all of the three high front types of articulation are counted as one because of the problem with the classification of [i]. Moreover, in terms of size-sound symbolism, the three varieties represent one class of sounds with a very high potential to symbolise "small size". For the same reason, no distinction is made between /ɪə/ and [iə].

The vowels which are presumed to symbolise "large size" are /ɑ:/ and /ɒ/. Both are clearly open and back. The interpretation of /ʌ/ and /ɔ:/, however, is more problematic. The former is an open centralised vowel, but it could be subsumed into the area of the lower-right hand corner of the vowel quadrilateral. The classification of /ɔ:/ is even less obvious. Although the vowel is clearly back, its pronunciation on the vertical articulatory scale is central. As a consequence, its potential to symbolise "large size" is presumed to be noticeably lower than those predicted for /ɑ:/ and /ɒ/.

The articulation of the central vowels /ɜ:/ and /ə/ is not expected to be related to any size distinctions. Except for the slightly more open realisation of /ʊ/ in the final position, they are pronounced in exactly the middle of both the vertical and horizontal articulatory scales. The size-sound symbolic potential of the other three remaining monophthongs – /æ/, /ʊ/ and /u:/ - is difficult to predict. The former is clearly an open vowel, but at the same time it is pronounced more to the front of the mouth than other open monophthongs. The latter two are articulated in the close-back region, which makes them "small" according to the horizontal scale, and "big" according to the vertical scale.

The RP diphthongs depicted in Figure 2 are also more difficult to interpret in terms of their size-sound symbolic potential. There are cases in which one part of the gliding vowel is supposedly "small" and the other one is "big", or vice versa. For this reason the diphthongs /aɪ/ and /ɔɪ/ are

presumed not to play a role in phonetic symbolism. Similarly, it is difficult to predict the exact potential of /eə/, /əʊ/ and /ʊə/. It could be speculated that /aʊ/ is rather "big", and /ɪʊ/ rather small, but the validity of this assumption is debatable. The only case which does not cause uncertainty is /eɪ/. According to the theories presented in the introduction, this diphthong should have the potential to symbolise objects that are small in size.

The analysis presented below is divided into two parts. In the first one the frequencies of the vowels found in the data are compared to the general frequencies of RP vowels suggested in Fry (1947). Such a comparison should reveal any universal strategies employed in the process of naming cars. The second part addresses the major aim of the paper and deals with the distribution of vowels in different groups of cars divided according to size.

3. Results and discussion

Table 1 presents a summary of the frequency of all the vowels found in the car names listed in the appendix. The percentage results have been calculated for the total 1478 instances of phonemes encountered in the analysed data. These results are compared with the frequency of RP vowels in transcribed spoken text proposed in Fry (1947).

Table 1. Comparison of the frequency of vowels found in the data and the results on general RP vowel frequencies reported in Fry (1947)

Expectations about vowel potential in sound symbolism	RP vowels	Frequencies of vowels in the current experiment	Text frequencies of vowels on the basis of Fry (1947)
"small"	/ɪ/, [i], /iː/	7.92%	9.98%
	/eɪ/	1.29%	1.71%
	/e/	2.50%	2.97%
unspecified	/ɪə/, [iə]	0.61%	0.21%
	/eə/	0.14%	0.34%
	/əʊ/	3.59%	1.51%
	/aɪ/	1.15%	1.83%
	/ɜː/	0.54%	0.52%
	/ə/	10.08%	10.74%
	/ʊ/	0.47%	0.86%
	/uː/	1.29%	1.13%
	/ʊə/	0.07%	0.06%
	/ɔɪ/	0.20%	0.14%
	/æ/	4.19%	1.45%
	/aʊ/	0.20%	0.61%
"big"	/ɔː/	1.08%	1.24%
	/ʌ/	0.74%	1.75%
	/ɒ/	1.76%	1.37%
	/ɑː/	2.44%	0.79%

It is readily visible that "small" vowels are generally less frequent in car names than in the summary reported in Fry (1947). The close front /ɪ/, [i] and /i:/ were observed in 7.92% of the cases in the current experiment, while Fry suggests that the frequency of their occurrence amounts to 9.98%. This difference is statistically significant ($p = 0.0082$). In the case of the other two "small" vowels the situation is analogous. Both /eɪ/ and /e/ were encountered less frequently than in Fry's analysis, but these differences cannot be statistically proven (in both cases $p > 0.05$).

The results for "big" vowels are less consistent. On the one hand, the "biggest" RP monophthongs /ɑ:/ and /ɒ/ were found to be more frequent in the current experiment than in Fry's publication. In the case of /ɑ:/ the difference must be regarded as highly statistically significant ($p < 0.0001$). The results for /ɒ/, however, are not statistically meaningful ($p = 0.1917$). Moreover, the frequency of occurrence for /ɔ:/ suggests that the vowel could actually be less frequent in car names than in Fry's analysis, but the difference is, again, statistically insignificant ($p = 0.5782$). Finally, /ʌ/ turned out to be less frequent in the current analysis than in Fry (1947) and the difference is statistically meaningful ($p = 0.0031$).

The data point to the conclusion that there is a general tendency of car manufacturers to create names with a relatively lower number of high front (or "small") vowels for their cars. Also, the producers tend to use a higher number of open back (or "big") vowels in naming their products, but this observation is less definite because of the contrary results for /ɔ:/. These tendencies may result from the fact that small cars are more desirable only in certain circumstances. For example, compact vehicles are practical for driving around the city and may be associated with lower petrol consumption. In general, however, large cars are more attractive to customers (at least those in the same price range as smaller models) as they tend to be perceived as more comfortable and providing ample luggage space. This is probably why companies prefer naming cars with the use of "bigger" rather than "smaller" vowels. Another plausible explanation for the tendency is connected with the fact that in the literature on sound-symbolism one may find dichotomies which are extensions of the basic contrast "big-small". For example, Jespersen (1922), Miron (1961) and Levickij (1971, 1973) suggest the opposition "strong-weak", Jespersen (1922) mentions "significant-insignificant" and Ohala (1994) discusses "dominant-subordinate". The extensions of the meanings "big" – "strong", "significant" and "dominant" – are features which are desirable in cars. Drivers want their vehicles to exhibit such characteristics. Conversely, the opposite associations – "weak", "insignificant" and "subordinate" –

should be avoided. Car manufactures want to communicate the positive qualities of their products to customers and this can also be done by the use of names with a relatively low number of high front vowels and a higher number of open back vowels.

As far as the frequency of occurrence of vowels which are unspecified for their role in sound symbolism is concerned, /əʊ/ and /æ/ are visibly more common in the current experiment than in Fry's summary, but such a distribution of the phonemes cannot be explained by any observations related to size-sound symbolism. In all other cases of the "unspecified" vowels the results obtained in the present study are similar to the ones reported in Fry (1947).

In order to examine the major problem raised in this paper – the possible influence of vehicle size on the distribution of vowels in corresponding car names – the 260 vehicles listed in the appendix have been divided into four groups of 85 cars. The first represents the smallest vehicles whose volume does not exceed 11054.7 dm^3. The second includes cars which are average. Their size ranges from 11163.7 dm^3 to 12784 dm^3. The third involves vehicles which are bigger than in the previous two groups. Their volume ranges from 12800.2 dm^3 to 15706.3 dm^3. Finally, the fourth group is comprised of the largest cars whose size is greater than 15706.3 dm^3. The frequency of occurrence of all the RP vowels has been calculated separately in each group and the results are presented in Table 2 below.

The only consistent correlation which can be observed concerns the high front /ɪ/, [i] and / i:/. The frequency of the vowels gradually decreases as the cars become bigger. It is the highest in the first group (10.37%), lower in the second (9.5%), still lower in the third (6.79%) and the lowest in the fourth (5.62%). More importantly, the results are statistically relevant. The p-value for the difference between the second and fourth groups equals 0.0413, and that for the difference between the first and fourth groups equals 0.0166. Consequently, it is a fact that car manufacturers tend to signal the size of the vehicles they produce by the frequency of high front vowels in car names.

The distribution of the other "small" vowels in Table 2 does not confirm the predicted tendencies. The frequency of their occurrence does not decrease with increasing size of the analysed vehicles. In the case of /eɪ/ the situation even seems to be reversed, since in the first group the diphthong appeared in 0.61% of the cases, and, for instance, in the third group in 2.09% of the cases. Still, this difference is not statistically relevant ($p = 0.0951$) and because the frequency in the fourth group decreases to 1.47%, no direct correlation can be confirmed.

Table 2. Comparison of the distribution of vowels in different groups of cars divided according to size

Expectations about vowel potential in sound symbolism	RP vowels	Group I	Group II	Group III	Group IV
"small"	/ɪ/, [i], /iː/	10.37%	9.50%	6.79%	5.62%
	/eɪ/	0.61%	0.84%	2.09%	1.47%
	/e/	2.13%	2.79%	2.35%	2.69%
unspecified	/ɪə/, [iə]	0.91%	0.28%	0.26%	0.98%
	/eə/	0.30%	0.28%	0.00%	0.00%
	/əʊ/	6.10%	3.35%	3.92%	1.47%
	/aɪ/	1.83%	0.56%	0.78%	1.47%
	/ɜː/	0.30%	0.28%	1.04%	0.49%
	/ə/	7.62%	10.89%	9.66%	11.74%
	/ʊ/	0.30%	0.28%	0.52%	0.73%
	/uː/	0.30%	2.79%	0.78%	1.22%
	/ʊə/	0.00%	0.00%	0.26%	0.00%
	/ɔɪ/	0.00%	0.00%	0.00%	0.73%
	/æ/	4.57%	3.07%	4.70%	4.40%
	/aʊ/	0.00%	0.00%	0.26%	0.49%
"big"	/ɔː/	1.22%	1.12%	0.78%	1.22%
	/ʌ/	0.00%	0.56%	0.52%	1.71%
	/ɒ/	1.52%	1.12%	3.13%	1.22%
	/ɑː/	2.13%	2.51%	1.83%	3.18%

Among the "big" vowels, only /ɑː/ and /ʌ/ behave in the expected way. Their frequency decreases with the increasing size of cars. Nevertheless, these tendencies are very weak and cannot be statistically proven. Also, the shifts in the distribution of /ɑː/ are not consistent because in the third group the vowel was, relatively speaking, the least frequent. Other vowels presumed to symbolise "large size" do not fall into patterns which could lead to any valid conclusions.

Among the vowels whose size-sound symbolic potential has not been specified, /əʊ/ behaves in a way which suggests correlation with the size of analysed cars. The larger the vehicles, the less frequent the occurrence of the vowel. Although this tendency is not fully consistent (the diphthong is slightly more common in the third group than in the second), it should be treated as statistically significant. The p-value for the difference between the frequency in the first group and the frequency in the fourth group equals 0.0007. Still, such a distribution of the vowel cannot really be explained in the sound symbolic framework adopted in this paper. From the purely

phonetic point of view, /əʊ/ begins with a quality similar to /ə/ and ends with an articulation similar to /ʊ/. None of these pronunciations can be clearly specified in terms of their potential to symbolise size. Moreover, the results in Table 2 do not support the possible claim that either /ə/ or /ʊ/ is suitable to represent small objects. In fact, in both cases the reverse could be speculated, since the frequency of the vowels tends to increase as the size of cars becomes larger.

All other cases of "unspecified" vowels do not reveal statistically significant results and no additional illations on the possible symbolic potential of these phonemes were drawn.

4. Conclusions

The analysis performed in this study indicates that size-sound symbolism is to a limited extent applied in car names. The frequency of high front vowels is the highest among names for small cars and it decreases with the increase in volume for the vehicles analysed. The reverse tendency among open back vowels, however, was not clearly confirmed. An additional observation made in the course of the analysis reveals that car manufacturers tend to make use of high front vowels in naming their products less frequently than can be observed in everyday speech. Conversely, open back vowels are utilised to a greater degree than was predicted in the summary prepared by Fry (1947). This implies that, in general, car names are designed in such a way as to be associated not only with the notion "big", but also with additional attributes such as "strong", "significant", "dominant", "heavy", etc. A possible future research could investigate these secondary associations. Some of them are directly measurable for cars and could be statistically evaluated in relation to sound symbolism.

REFERENCES

Åsa, Abelin
1999 *Studies in Sound Symbolism.* (Gotheburg Monograms in Linguistics 17.) Göteborg: Göteborg University Press.

Bentley, Madison – Edith J. Varon
1933 "An accessory study of phonetic symbolism", *American Journal of Psychology* 45, 76-86.

Bolinger, Dwight
1950 "Rime, assonance and morpheme analysis", *Word* 6, 117-136.
Bolinger, Dwight
1965 *Forms of English: Accent, Morpheme, Order.* Tokyo: Hokuou Publishing Company.
Brackbill, Yvonne – Kenneth B. Little
1957 "Factors determining the guessing of meanings of foreign words", *Journal of Abnormal and Social Psychology* 54, 312-318.
Brown, Roger W. et al.
1955 "Phonetic symbolism in natural languages", *Journal of Abnormal and Social Psychology* 50, 388-393.
Caltvedt, Les
1999 "The poetics of synaesthesia in Rilke and Handke", *Modern Austrian Literature* 32 (1), 53-66.
Chapman, Raymond
1982 *The Language of English Literature.* London: Edward Arnold Publishers Ltd.
Frazer, June M.
1982 "The phonetic symbolism of Arnold's 'Wragg is in custody'", *American Notes & Queries* 21, 13.
Fry, Dennis B.
1947 "The frequency of occurrence of speech sounds in southern English", *Archives Néerlandaises de Phonétique Expérimentale* 20, 103-106.
Gimson, Alfred C.
1994 *Gimson's Pronunciation of English.* London: Edward Arnold Publishers Ltd.
Hinton, Leanne et al. (eds.)
1994 *Sound Symbolism*. Cambridge: Cambridge University Press.
Householder, Fred W.
1960 "From the viewpoint of linguistics". In: T.A. Sebeok (ed.) *Style in Language.* Cambridge, Massachusetts: The M.I.T. Press, 339-349.
Hymes, Dell H.
1960 "Phonological aspects of style: some English sonnets". In: T.A. Sebeok (ed.) *Style in Language.* Cambridge, Massachusetts: The M.I.T. Press, 109-131.
Jakobson, Roman – Linda R. Waugh
1979 *The Sound Shape of Language.* Sussex: Harvester Press.
Jespersen, Otto
1922 *Language: Its Nature, Development and Origin.* London: Allen and Unwin.
Klink, Richard R.
2000 "Creating brand names with meaning: the use of sound symbolism", *Marketing Letters* 11 (1), 5-20.

Levickij, Viktor V.
1971 "Čy isnuje universalij zvukosymvolizm?", *Movoznavstvo* 1, 25-32.
Levickij, Viktor V.
1973 *Semantika i Fonetika. Posobie, Podgotovlennoe na Materiale Eksperimentalnyx Issledovanij.* Černovcy: Čenrovickij Gosudarstvennyj Universitet.
Lowrey, Tina M. – L. J. Shrum
2007 "Phonetic symbolism and brand name preference", *Journal of Consumer Research* 34 (3), 406-414.
Lucas, Frank L.
1955 *Style.* London: Cassell & Co Ltd.
Malkiel, Yakov
1994 "Regular sound development, phonosymbolic orchestration, disambiguation of homonyms". In: L. Hinton et al. (eds.) *Sound Symbolism.* Cambridge: Cambridge University Press, 207-221.
Maltzman, Irving et al.
1956 "An investigation of phonetic symbolism", *Journal of Abnormal and Social Psychology* 53, 249-251.
Marchand, Hans
1960 *The Categories and Types of Present-Day English Word-Formation.* Wiesbaden: Otto Harrassowitz.
Markell, Norman N. – Eric P. Hamp
1960 "Connotative meanings of certain phoneme sequences", *Studies in Linguistics* 15, 47-61.
McCune, Keith M.
1985 *The Internal Structure of Indonesian Roots.* Atma Jaya: Universitas Katolik Indonesia.
Murdy, Louise B.
1966 *Sound and Sense in Dylan Thomas's Poetry.* The Hague: Mouton & Co.
Miron, Murray S.
1961 "A cross-linguistic investigation of phonetic symbolism", *Journal of Abnormal and Social Psychology* 62 (3), 623-630.
Nash, Walter
1980 *Designs in Prose.* London: Longman.
Newman, Stanley S.
1933 "Further experiments in phonetic symbolism", *American Journal of Psychology* 45, 53-75.
Nichols, Johanna
1971 "Diminutive consonant symbolism in Western North America", *Language* 47 (4), 826-848.
Nordberg, Bengt
1986 *The Use of Onomatopoeia in the Conversational Style of Adolescents.* (FUMS rapport nr 132.) Uppsala: Uppsala University (FUMS).

Nuckolls, Janis B.
1999 "The case for sound symbolism", *Annual Review of Anthropology* 28, 225-252.
Ohala, John J.
1983 "Cross-language use of pitch: an ethological view", *Phonetica* 40, 1-18.
1984 "An ethological perspective on common cross-language utilization of F_0 of voice", *Phonetica* 41, 1-16.
1994 "The frequency code underlies the sound-symbolic use of voice pitch". In: L.Hinton et al. (eds.) *Sound Symbolism*. Cambridge: Cambridge University Press, 325-347.
Rhodes, Richard
1994 "Aural images". In: L. Hinton et al. (eds.) *Sound Symbolism*. Cambridge: Cambridge University Press, 276- 292.
Roach, Peter
2009 *English Phonetics and Phonology.* Cambridge: Cambridge University Press.
Sapir, Edward
1929 "A study in phonetic symbolism", *Journal of Experimental Psychology* 12, 225-239.
Stolarski, Łukasz
Forthcoming "Palatality in size-sound symbolism on the basis of selected Polish examples". In: M. Mochoń (ed.) *Studia Neofilologiczne.* Kielce: Wydawnictwo Uniwersytetu Jana Kochanowskiego w Kielcach.
Tolman, Albert H.
1906 "The symbolic value of English sounds". In: A.H. Tolman (ed.) *The Views about Hamlet (and other Essays).* Boston: Houghton, Mifflin and Company, l43-172.
Ultan, Russell
1978 "Size-sound symbolism". In: J.H. Greenberg et al. (eds.) *Universals of Human Language.* Stanford, California: Stanford University Press, 525-568. Urban, Matthias
2011 "Conventional sound symbolism in terms for organs of speech: A cross-linguistic study", *Folia Linguistica* 45 (1), 199-213.
Wells, John C.
2008 *Longman Pronunciation Dictionary.* Harlow: Person Longman.
Wichmann, Soren et al.
2010 "Sound symbolism in basic vocabulary", *Entropy* 12, 844-858.
Yorkston, Eric – Geeta Menon
2004 "A sound idea: phonetic effects of brand names on consumer judgements", *Journal of Consumer Research* 31 (1), 43-51.

APPENDIX

Model	Predicted RP Pronunciation	Length (mm)	Width (mm)	Height (mm)	Overall size (dm^3)
GMC					
Acadia 2011	/əˈkeɪdiə/	5110	2010	1860	19104,2
Canyon Crew Cab 4x4 SLE-1 2011	/ˈkænjən/	5270	1750	1650	15217,1
Savana Cargo Van G1500 2011	/səˈvænə/	5700	2020	2130	24524,8
Sierra 1500 2011	/siˈerə/	5230	2040	1880	20058,1
Terrain 2011	/təˈreɪn/	4710	1860	1690	14805,4
Yukon 2011	/ˈjuːkɒn/	5140	2010	1960	20249,5
Envoy 2009	/ˈenvɔɪ/	4870	1900	1830	16933,0
Honda					
Accord 2.0 Sport 2011	/əˈkɔːd/	4670	1770	1450	11985,6
Civic 1.3 i-DSi VTEC Hybrid 2010	/ˈsɪvɪk/	4550	1760	1440	11531,5
Element 2011	/ˈelɪmənt/	4320	1790	1830	14151,0
Fit 2011	/fɪt/	4110	1700	1530	10690,1
Insight 1.3 2009	/ˈɪnsaɪt/	4400	1700	1430	10696,4
Jazz 1.2 i-VTEC 2011	/dʒæz/	3910	1700	1530	10169,9
Legend 3.5i V6 2011	/ˈledʒənd/	4990	1820	1440	13077,8
Odyssey EX 2011	/ˈɒdəsi/	5160	2020	1740	18136,4
Ridgeline 2011	/ˈrɪdʒlaɪn/	5260	1980	1790	18642,5
Pilot 2010	/ˈpaɪlət/	4860	2780	1810	24454,5
Stream 1.7i ES 2009	/striːm/	4580	1700	1600	12457,6
Hyunday					
Accent 2011	/ˈæksənt/	4290	1700	1480	10793,6
Equus Ultimate 2011	/ˈekwəs/	5170	1900	1500	14734,5
Genesis Coupe 2.0T 2010	/ˈdʒenəsɪs/	4640	1870	1390	12060,8
Getz 1.1 GL 2011	/gets/	3820	1670	1500	9569,1
Santa Fe 2.2 CRDi 2011	/sæntəˈfeɪ/	4670	1900	1770	15705,2
Sonata 2.0 GLS 2011	/səˈnɑːtə/	4750	1830	1430	12430,3
Terracan 2.9 CRDi GL 2011	/ˈterəkæn/	4720	1870	1800	15887,5
Tucson 2.0 4WD GLS 2011	/ˈtuːsɒn/	4330	1800	1740	13561,6
Veracruz 2011	/vɪərəˈkruːz/	4850	1950	1810	17118,1
Grandeur 2.2 CRDi 2009	/ˈgrændʒə/	4900	1870	1500	13744,5
Matrix 1.6 2010	/ˈmeɪtrɪks/	4060	1750	1640	11652,2
Ceed 1.4 CVVT 2011	/siːd/	4270	1800	1490	11452,1
Forte EX 2011	/ˈfɔːteɪ/	4490	1770	1410	11205,7
Optima EX Turbo 2011	/ˈɒptɪmə/	4850	1840	1460	13029,0
Picanto 1.1 2011	/pɪˈkæntəʊ/	3540	1600	1490	8439,4
Rio 1.4 2010	/ˈriːəʊ/	4000	1700	1480	10064,0
Sorento 2.2 CRDi 2011	/səˈrentəʊ/	4690	1890	1760	15600,8
Soul 1.6 2011	/səʊl/	4110	1790	1620	11918,2
Sportage 2.0 2010	/ˈspɔːtɑːʒ/	4360	1810	1700	13415,7
Venga 1.4 CVVT 2011	/ˈvengə/	4070	1770	1610	11598,3
Cerato 1.6 LX 2009	/səˈrɑːtəʊ/	4350	1740	1480	11202,1
Magentis 2.0 SE 2008	/məˈdʒentɪs/	4740	1830	1420	12317,4
Rondo 2.4 2007	/ˈrɒndəʊ/	4550	1830	1660	13822,0
Spectra 2.0 2008	/ˈspektrə/	4510	1740	1480	11614,2
Carnival 2.7 V6 EX 2007	/ˈkɑːnɪvl/	4820	1990	1820	17457,1
Mitsubishi					
Colt 1.1 2011	/kəʊlt/	3890	1700	1530	10117,9
Eclipse 2010	/ɪˈklɪps/	4590	1840	1370	11570,5
Endeavor 2011	/ɪnˈdevə/	4850	1880	1770	16138,9
Lancer 1.5 2011	/ˈlɑːnsə/	4580	1770	1500	12159,9
Outlander 2.0 2011	/ˈaʊtlændə/	4670	1810	1690	14285,1
Grandis 2.0 Di-D 2008	/ˈgrændɪs/	4770	1800	1660	14252,8
Space Star 1.3 Family 2008	/ˈspeɪs stɑː/	4060	1720	1520	10614,5
Nissan					
Altima 2.5 2011	/ˈɔːltɪmə/	4850	1800	1480	12920,4
Armada Platinum 2011	/ɑːˈmɑːdə/	5280	2020	1970	21011,2
Cube 1.5 dCi 2010	/kjuːb/	3990	1700	1680	11395,4
Frontier King Cab S 2011	/ˈfrʌntɪə/	5230	1860	1750	17023,7
Juke 1.5 dCi 2011	/dʒuːk/	4140	1760	1580	11512,5
Maxima QX 3.0 2011	/ˈmæksɪmə/	4930	1790	1440	12707,6
Micra 1.2 2011	/ˈmaɪkrə/	3790	1670	1530	9683,8
Note 1.4 2010	/nəʊt/	4110	1700	1560	10899,7
Pathfinder 2.5 dCi 2011	/ˈpɑːθ faɪndə/	4750	1860	1790	15814,7
Pixo 1.0 2011	/ˈpɪksəʊ/	3570	1610	1480	8506,6
Quest 3.5 2008	/ˈkwest/	5190	1980	1830	18805,4
Rogue 2011	/ˈrəʊg/	4660	1810	1670	14085,8
Sentra 2.0 2011	/ˈsentrə/	4570	1800	1520	12503,5
Titan 2011	/ˈtaɪtn/	5710	2030	1900	22023,5
Versa 1.6 2011	/ˈvɜːsə/	4480	1700	1540	11728,6
Patrol 3.0 TD GL 2009	/pəˈtrəʊl/	5040	1950	1860	18280,1
Hardbody 2400i 4x4 2007	/ˈhɑːdbɒdi/	4910	1830	1680	15095,3
Qashqai 1.5 dCi 2011	/ˈkæʃkaɪ/	4550	1790	1650	13438,4
Murano 3.5 V6 2011	/məˈrɑːnəʊ//	4870	1890	1730	15923,4
Navara 2.5 dCi 2010	/nəˈvɑːrə/	5140	1860	1760	16826,3

Model	Predicted RP Pronunciation	Length (mm)	Width (mm)	Height (mm)	Overall size (dm^3)
Renault					
Clio 1.2 2011	/ˈkliːəʊ/	4040	1730	1500	10483,8
Espace 1.9 dCi Avantage 2008	/ɪˈspɛɪs/	4670	1900	1700	15084,1
Fluence 1.5 dCi 110 FAP Eco 2011	/ˈfluːənz/	4620	1820	1480	12444,4
Kangoo 1.2 Campus 2011	/kæŋˈguː/	4040	1680	1830	12420,6
Laguna 1.5 dCi 110 FAP 2011	/ləˈguːnə/	4700	1820	1450	12403,3
Latitude 2.5 2011	/ˈlætɪtjuːd/	4890	1840	1490	13406,4
Megane 1.4 Authentique 2011	/məˈgæn/	4220	1780	1460	10966,9
Modus 1.2 2011	/ˈməʊdəs/	3880	1720	1600	10677,8
Scenic 1.4 Authentique 2008	/ˈsiːnɪk/	4270	1820	1630	12667,4
Twingo 1.2 Authentique 2008	/ˈtwɪŋgəʊ/	3440	1640	1430	8067,5
Trafic 1.9 DCi Van 2007	/ˈtræfɪk/	4790	2240	1970	21137,3
Seat					
Alhambra 1.4 TSi 2011	/ælˈhæmbrə/	4860	1910	1730	5716,4
Ibiza 1.2 2011	/ɪˈbiːθə/	4060	1700	1450	10007,9
Leon 1.4 2011	/ˈliːən/	4320	1770	1460	11163,7
Cordoba 1.2 Reference 2008	/ˈkɔːdəbə/	4290	1700	1450	10574,9
Toledo 1.6 Reference 2008	/təˈliːdəʊ/	4450	1750	1440	11214,0
Suzuki					
Alto 1.1 Classic 2011	/ˈæltəʊ/	3500	1480	1460	7562,8
Equator 2011	/ɪˈkweɪtə/	5250	1860	1750	17088,8
Grand Vitara 1.6 2011	/ˈgrænd vɪˈtɑːrə/	3870	1820	1690	11903,3
Jimny 1.3 2011	/ˈdʒɪmni/	3670	1610	1680	9926,6
Kizashi Sport 2011	/kəˈzɑːʃi/	4660	1830	1490	12706,4
Splash 1.0 2011	/splæʃ/	3720	1690	1600	10058,9
Swift 1.2 DDiS 2011	/swɪft/	3860	1700	1520	9974,2
Ignis 1.3 DDiS Club 2008	/ˈɪgnɪs/	3780	1610	1570	9554,7
Forenza 2007	/fəˈrenzə/	4510	1730	1450	11313,3
Liana 1.3 Club 2008	/liːˈɑːnə/	4360	1700	1550	11488,6
Reno Convenience 2008	/ˈriːnəʊ/	4300	1730	1450	10786,6
Aerio AWD 2007	/ˈeərɪəʊ/	4360	1730	1550	11691,3
Verona 2006	/vəˈrəʊnə/	4780	1820	1460	12701,4
Toyota					
4Runner Limited V6 2011	/ˈfɔːrʌnə/	4830	1930	1820	16965,9
Auris 1.33 2010	/ˈɔːrɪs/	4250	1770	1520	11434,2
Avalon 3.5 2010	/ˈævəlɒn/	5030	1860	1490	13940,1
Aygo 1.0 2011	/ˈeɪgəʊ/	3420	1620	1470	8144,4
Camry Hybrid 2010	/ˈkæmri/	4810	1830	1470	12939,4
Corolla 1.3 Advanced 2010	/kəˈrɒlə/	4550	1770	1480	11919,2
FJ Cruiser 2010	/ˈkruːzə/	4680	1910	1820	16268,6
Highlander 2.7 2010	/ˈhaɪləndə/	4790	1920	1740	16002,4
Hilux 2.5 D-4D SingleCab 2010	/ˈhaɪlʌks/	5260	1770	1690	15734,2
Land Cruiser 3.0 D-4D 4WD 2011	/ˈlænd kruːzə/	4770	1890	1900	17129,1
Prius 2011	/ˈpraɪəs/	4470	1750	1500	11733,8
Sequoia 2011	/sɪˈkwɔɪə/	5220	2040	1900	20232,7
Sienna 2011	/siˈenə/	5090	1990	1760	17827,2
Tacoma 2010	/təˈkəʊmə/	4840	1840	1680	14961,4
Tundra Regular Cab 4.0L V6 2011	/ˈtʌndrə/	5340	2040	1930	21024,6
Urban Cruiser 1.33 2011	/ˈɜːbən/	3940	1730	1530	10428,8
Venza 2010	/ˈvenzə/	4810	1910	1620	14883,1
Verso 1.6 2011	/ˈvɜːsəʊ/	4450	1800	1630	13056,3
Yaris 1.0 2011	/ˈjɑːrɪs/	3790	1700	1540	9922,2
Fortuner 3.0D-4D Automatic 2010	/ˈfɔːtʃənə/	4700	1850	1860	16172,7
Sprinter 140i 2008	/ˈsprɪntə/	4180	1700	1480	10516,9
Quantum 2.5 D4-D Bus 2007	/ˈkwɒntəm/	5390	1890	2290	23328,5
Volkswagen					
Beetle 1.4 2011	/biːtl/	4140	1730	1500	10743,3
Caddy 1.2 TSI 2011	/ˈkædi/	4410	1800	1860	14764,7
Eos 1.4 TSi 2011	/ˈiːɒs/	4410	1800	1450	11510,1
Fox 1.2 2011	/fɒks/	3830	1670	1550	4012,2
Golf 1.2 TSi 2011	/gɒlf/	4210	1790	1490	11228,5
Jetta 1.2 TSI 2011	/ˈdʒetə/	4650	1780	1460	12084,4
Multivan 1.9 TDi Startline 2011	/ˈmʌltivæn/	4900	1910	1950	18250,1
Passat 1.4 TSi 2011	/pæˈsæt/	4780	2070	1480	14644,0
Phaeton 3.0 V6 TDi 4Motion 2011	/ˈfeɪtn/	5060	1910	1460	14110,3
Polo 1.2 Fun 2011	/ˈpəʊləʊ/	3910	1680	1530	10050,3
Scirocco 1.4 TSI 2011	/sɪˈrɒkəʊ/	4260	1820	1410	10932,0
Sharan 1.8 T 2010	/ˈʃærən/	4640	1820	1740	14694,0
Touareg V6 2011	/tʊˈɑːreg/	4800	1950	1710	16005,6
Citi Sport 1.4i 2009	/ˈsɪti spɔːt/	3820	1400	1620	8663,8
Kombi 1.9 TDi 2009	/ˈkɒmbi/	4900	1970	1910	18437,2
Rabbit 2007	/ˈræbɪt/	4220	1770	1480	11054,7
Bora 1.4 2008	/ˈbɔːrə/	4380	1740	1450	11050,7
Amarok 2.0 TDi 4x4 2011	/ˈæmərɒk/	5260	1950	1840	18872,9
Routan 2011	/ruːˈtɑːn/	5150	1960	1760	17765,4

Model	Predicted RP Pronunciation	Length (mm)	Width (mm)	Height (mm)	Overall size (dm^3)
Alfa Romeo					
Giulietta 1.4 TB 2011	/dʒuːliˈetə/	4360	1800	1470	11536,6
MiTo 1.3 JTDM 2011	/ˈmiːtəʊ/	4070	1730	1460	10280,0
Spider 2.0 JTS 2011	/ˈspaɪdə/	4310	1780	1320	10126,8
Brera 2.2 JTS 16V 2009	/ˈbrerə/	4420	1840	1350	10979,3
Aston Martin					
Cygnet 2011	/ˈsɪgnət/	3078	1680	1500	7756,6
DB9 Volante 2011	/vəˈlænti/	4720	1880	1270	11269,5
Rapide 2011	/rəˈpiːd/	5030	2150	1330	14383,3
Vantage V8 2011	/ˈvɑːntɪdʒ/	4390	1870	1260	10343,7
Virage Coupe 2011	/ˈvɪrɑːʒ/	4710	1910	1290	11605,0
Vanquish S V12 2009	/ˈvæŋkwɪʃ/	4670	1930	1340	12077,6
Bently					
Continental GT 2011	/kɒntɪˈnentl/	4810	1920	1400	12929,3
Mulsanne 2011	/mʊlˈsɑːn/	5580	1930	1530	16477,2
Azure 2009	/əˈzjʊə/	5420	1910	1500	15528,3
Arnage T 2009	/əˈnɑːʒ/	5410	1910	1520	15706,3
Brooklands 2009	/ˈbrʊklənds/	5410	1910	1520	15706,3
Buick					
Enclave 2011	/ˈenkleɪv/	5120	2010	1850	19038,7
LaCrosse 2010	/ləˈkrɒs/	5010	1860	1500	13977,9
Lucerne 2011	/luˈsɜːn/	5170	1880	1480	14385,0
Regal GS 2010	/ˈriːgl/	4831	1811	1473	12887,2
RAINIER CXL 4.2L (2007)	/ˈreɪniə/	4912	1915	1826	17176,2
Chevrolet					
Avalanche 2011	/ˈævəlɑːntʃ/	5630	2020	1950	22176,6
Aveo 1.2 2010	/ˈæviəʊ/	3930	1690	1510	10029,0
Camaro 2010	/kəˈmærəʊ/	4840	1920	1380	12824,1
Captiva 2.0 D 2011	/kæpˈtiːvə/	4640	1860	1730	14930,6
Colorado 2011	/kɒləˈrɑːdəʊ/	4890	1720	1650	13877,8
Corvette 2011	/kɔːˈvet/	4440	1850	1250	10267,5
Cruze 1.6 2011	/kruːz/	4600	1790	1480	12186,3
Equinox 2011	/ˈekwɪnɒks/	4780	1850	1690	14944,7
Impala 2011	/ɪmˈpɑːlə/	5100	1860	1500	14229,0
Malibu 2011	/ˈmælɪbuː/	4880	1790	1460	12753,4
Silverado 1500 2011	/sɪlvəˈrɑːdəʊ/	5230	2040	1880	20058,1
Spark 1.2 2010	/spɑːk/	3650	1600	1530	8935,2
Suburban 2011	/səˈbɜːbən/	5660	2020	1960	22409,1
Tahoe 5.3 2011	/ˈtɑːhəʊ/	5140	2010	1960	20249,5
Traverse 2011	/ˈtrævɜːs/	5210	2000	1860	19381,2
Volt 2011	/vəʊlt/	4500	1790	1440	11599,2
Cobalt 1LT 2010	/ˈkəʊbɔːlt/	4590	1730	1460	11593,4
Epica 2.0 2010	/ˈepɪkə/	4810	1820	1460	12781,1
Lumina SS 6.0 UTE 2010	/ˈluːmɪnə/	5060	1480	1850	13854,3
Trail Blazer 4.2 LT 2010	/ˈtreɪlbleɪzə/	4900	1910	1830	17127,0
Optra 1.6 L 2009	/ˈɒptrə/	4510	1730	1450	11313,3
Trans Sport 3.4 2008	/ˈtrænspɔːt/	5120	1850	1810	17144,3
Uplander Cargo Van 2008	/ˈʌplændə/	5200	1840	1840	17605,1
Monte Carlo 2006	/mɒntɪˈkɑːləʊ/	5000	1860	1420	13206,0
Chrysler					
Voyager 2.4 Family 2010	/ˈvɔɪɪdʒə/	4810	2000	1760	16931,2
PT Cruiser 1.6 Classic 2009	/ˈkruːzə/	4290	1750	1610	12087,1
Sebring LX 2.0 2010	/ˈsiːbrɪŋ/	4850	1800	1400	12222,0
Town & Country 2010	/ˈtaʊnənˈkʌntri/	5150	1960	1760	17765,4
Aspen 2008	/ˈæspən/	5140	1940	1880	18746,6
Crossfire 3.2 Roadster V6 LTD	/ˈkrɒsfaɪə/	4060	1770	1320	9485,8
Pacifica 2007	/pəˈsɪfɪkə/	5050	2020	1740	17749,7
Citroen					
Berlingo 1.4i 2007	/bɜːˈlɪŋgəʊ/	4110	1730	1810	12869,6
Nemo 1.4 2011	/ˈniːməʊ/	3970	2030	1740	14022,8
Xsara 1.4 HDi SX Plus 2011	/ˈzærə/	4190	1710	1410	10102,5
Picasso 1.6i HDi Exclusive 2008	/pɪˈkæsəʊ/	4280	1640	1760	12353,8
Multispace 1.4 2008	/ˈmʌltispeɪs/	4140	1730	1820	13035,2
Dacia					
Duster 1.6 2011	/ˈdʌstə/	4320	1830	1630	12886,1
Logan MCV 1.4 2011	/ˈləʊgən/	4460	1750	1640	12800,2
Sandero 1.2 Eco 2011	/sænˈdeərəʊ/	4030	1750	1540	10860,9

Model	Predicted RP Pronunciation	Length (mm)	Width (mm)	Height (mm)	Overall size (dm^3)
Daewoo					
Matiz 0.8 S 2010	/məˈtiːz/	3500	1500	1490	7822,5
Nubira 1.6 SE 2010	/nʊˈbɪrə/	4510	1730	1450	11313,3
Rezzo 1.6 SX 2010	/ˈretsəʊ/	4360	1760	1590	12201,0
Evanda 2.0 CDX 2008	/ɪˈvændə/	4780	1820	1450	12614,4
Daihatsu					
Copen 0.7 2011	/ˈkəʊpən/	3400	1480	1250	6290,0
Materia 1.3 2010	/məˈtɪərɪə/	3810	1700	1640	10622,3
Sirion 1.0 2010	/ˈsɪrɪən/	3610	1670	1560	9404,8
Terios 1.3 2008	/ˈterɪɒs/	3850	1560	1720	10330,3
Charade CX 2007	/ʃəˈrɑːd/	3410	1480	1510	7620,7
Trevis 2009	/ˈtrevɪs/	3420	1500	1430	7335,9
Dodge					
Avenger 2011	/əˈvendʒə/	4900	1860	1490	13579,9
Caliber 1.8 L SXT 2009	/ˈkælɪbə/	4420	1750	1540	11911,9
Challenger 2009	/ˈtʃælɪndʒə/	5030	1930	1460	14173,5
Charger 2010	/ˈtʃɑːdʒə/	5090	1900	1480	14313,1
Dakota 2011	/dəˈkəʊtə/	5560	1830	1750	17805,9
Durango Crew 2011	/djʊˈræŋgəʊ/	5080	1930	1810	17746,0
Journey 2.0 2011	/ˈdʒɜːni/	4890	1880	1700	15628,4
Nitro 2.8 CRD 2010	/ˈnaɪtrəʊ/	4590	1860	1780	15196,6
Ram 1500 2010	/ræm/	5320	2020	1900	20418,2
Viper SRT-10 2011	/ˈvaɪpə/	4470	1950	1220	10634,1
Magnum 2008	/ˈmægnəm/	5030	1890	1490	14165,0
Caravan 2007	/ˈkærəvæn/	4810	2000	1760	16931,2
Stratus 2005	/ˈstreɪtəs/	4850	1800	1400	12222,0
Fiat					
Bravo 1.4 2011	/ˈbrɑːvəʊ/	4340	1800	1500	11718,0
Doblo 1.2 Trofeo 2011	/ˈdɒbləʊ/	4170	1720	1820	13053,8
Panda 1.1 2011	/ˈpændə/	3540	1600	1550	8779,2
Punto 1.2 2008	/ˈpʊntəʊ/	3850	1670	1490	9580,0
Qubo 1.3 Multijet 2011	/ˈkjuːbəʊ/	3970	1720	1740	11881,4
Strada 1.3 Multijet 2010	/ˈstrɑːdə/	4450	1670	1600	11890,4
Croma 1.8 2009	/ˈkrəʊmə/	4790	1780	1610	13727,2
Idea 1.2 Active 2009	/aɪˈdɪə/	3940	1700	1670	11185,7
Linea 1.3 Multijet 2009	/ˈlɪnɪə/	4570	1950	1500	13367,3
Multipla 1.6 Active 2009	/ˈmʌltɪplə/	4000	1880	1700	12784,0
Palio II 1.2 EL 2007	/ˈpælɪəʊ/	3830	1640	1440	9044,9
Siena II 1.2 EL 2007	/ˈsiˈenə/	3830	1640	1440	9044,9
Ford					
Edge 2011	/edʒ/	4720	1930	1710	15577,4
Escape 2011	/ɪˈskeɪp/	4440	1810	1730	13903,0
Expedition 2011	/ekspəˈdɪʃn/	5250	2010	1970	20788,4
Explorer 2010	/ɪkˈsplɔːrə/	4920	1880	1860	17204,3
Fiesta 1.25 2011	/fiˈestə/	3960	1730	1490	10207,7
Flex 2011	/fleks/	5130	2260	1730	20057,3
Focus 1.4 2011	/ˈfəʊkəs/	4340	1890	1500	12303,9
Fusion 1.25 2011	/ˈfjuːʒn/	4020	1730	1500	10431,9
Galaxy 2.0 2011	/ˈgæləksi/	4830	2160	1810	18883,4
Ka 1.2 2011	/kɑː/	3630	1660	1510	9099,0
Maverick 2.0i Highclass 2011	/ˈmævərɪk/	4420	1830	1780	14397,7
Mondeo 1.6 Ti-VCT 2010	/mɒnˈdeɪəʊ/	4850	2080	1510	15232,9
Mustang 2010	/ˈmʌstæŋ/	4770	1880	1410	12644,3
Ranger 2.3 2011	/ˈreɪndʒə/	4820	1770	1690	14418,1
Taurus 2011	/ˈtɔːrəs/	5160	1940	1550	15516,1
Transit Connect 2011	/ˈtrænzɪt/	1390	4590	1800	11484,2
Bantam 1.3i 2009	/ˈbæntəm/	4280	1470	1640	10318,2
Shelby GT 500 2009	/ˈʃelbi/	4770	1890	1390	12531,3
Streetka 1.6 2009	/ˈstriːtkɑː/	3660	1690	1350	8350,3
Ikon 1.3i L 2007	/ˈaɪkɒn/	4150	1640	1390	9460,3
Freestar 2007	/ˈfriːstɑː/	5110	1950	1800	17936,1
Freestyle 2006	/ˈfriːstaɪl/	5100	1900	1530	14825,7
Territory 4.0 Ghia Automatic 2007	/ˈterətri/	4860	1720	1900	15882,5
Van E-150 2007	/ˈvæn/	5390	2020	2060	22428,9
Wagon E-150 Chateau 2007	/ˈwægən/	5390	2020	2060	22428,9
Crown Victoria 2006	/ˈkraʊn vɪkˈtɔːriə/	5390	2000	1490	16062,2


Druk i oprawa
Wydawnictwo UJK

Token: A Journal of English Linguistics focuses on English linguistics in a broad sense, and accepts both diachronic and synchronic work, grammatical as well as lexical studies. That being said, the journal favors empirical research. Jan Kochanowski University (Kielce, Poland) publishes *Token* once annually, and all submissions are double-blind peer reviewed. The journal's website is at http://www.ujk.edu.pl/token/.