A statistical analysis of the "Eumaeus" Phrasemes in James Joyce's *Ulysses*

Robert Byrnes

Department of Rhetoric, Writing and Communications
University of Winnipeg, MB R3B 2E9 – Canada

Abstract

In his 1922 modernist masterpiece *Ulysses*, James Joyce displays his stylistic brilliance by using a new, different style in each of the last 8 of the book's 18 chapters (which he called episodes). In the 16th episode, unofficially entitled "Eumaeus," Joyce's narrative is left to his main character, Leopold Bloom, whose long harangue to Stephen Daedalus, very late into the night, is strewn with clichéd idioms. Several critics have identified a few handfuls of them, and one estimated there might be as many as 700, but none has tried to identify them all, and none was able to fully explain why Joyce himself described the "Eumaeus" narrative as "rilassatta," or "tired" prose. Close reading reveals the extraordinary extent of Joyce's cliché project. Indeed, there is hardly a phrase in "Eumaeus" that isn't a cliché. "Eumaeus" is a thesaurus of Dublin cliché in 1904, and Joyce was a phraseologist *avant la lettre*. My phraseological analysis led me to construct a database (unpublished) of at least 2561 "Eumaeus" clichés, along with fields for Lexical, Semantic, Stylistic, etc., attributes for each cliché. The Semantic and Stylistic analysis is crucial for translators, because Joyce's textual drama depends upon the collision of disparate Semantic and Stylistic levels as the discourse moves from cliché to cliché. Bloom is in fact a linguistically incompetent autodidact striving for stylistic effects above his native register. This makes "Eumaeus" the most difficult chapter for translators in all of English literature, and more so because such a large percentage of the clichés are transparent collocations rather than fixed, opaque, and easily identifiable traditional idioms. This paper presents my textual database of "Eumaeus" handy for accurate translation, and a variety of statistical results useful for critical interpretation. The results make it clear that only a statistical database can adequately describe crucial dimensions of Joyce's phraseological art.

Keywords: phraseology, translation, cliché, lexical type, semantic type, style level, James Joyce, *Ulysses*

1. Introduction

Ulysses, by James Joyce, is the notoriously demanding Mount Everest of the English novel, a modernist masterpiece that begins with a demanding new technique, stream of consciousness, and then becomes even more demanding when its style changes drastically in each of the last 8 of its 18 episodes. It has spawned a "Joyce industry" of annual conventions national and international, Joyce journals, and a yearly holiday, June 16th, the single day on which the novel takes place, celebrated across every reasonably literate country in the world with pub crawling and readings from "Penelope," the book's final episode, in which an unfaithful wife remembers her infidelity of that very afternoon, while her husband sleeps beside her. And it's a record without peer of Dublin English on the day the action takes place, with every level of speech, dialect, accent, piece of slang or vulgarity available in the speech of the time, from high-style political rodomontade in the newspaper offices to barfly invective in a local public house. It's a stunning encyclopedia and dictionary of Irish English on July 16, 1904, and it includes within

it a canvassing of virtually every written genre, from sports journalism to fashion magazine features, from funeral notices to poetry. One of its episodes, "Circe", is simply an expressionist drama Joyce bootlegged into his novel.

The episode I've created a database for is one of the stylistic experiments in the latter part of the novel, in third-to-last position, at the beginning of the *Nostos* or *Return* section of the novel, when Ulysses finally reaches his home island after a 20-year absence. It's unofficially called "Eumaeus" and it's named after the pigherd in the *Odyssey*, who shelters Ulysses in his hut without recognizing his former master. Joyce set himself the task of writing it entirely in the everyday language of Dublin in 1904, and in the voice of one of its uneducated denizens, Leopold Bloom, the Ulysses figure of the novel. More than 99% is simply Bloom's voluble talk to his drunk and silent auditor Stephen Daedalus, as they sit in a cabman's shelter sipping coffee. Bloom wishes to impress Stephen, who has just graduated with a B. A. in modern languages, so he pours out an endless stream of commentary on a miscellany of topics, but all in the phrasing of an unsophisticated autodidact. This monologue has always been understood to include a significant number of clichés, but each reader has been left to surmise how many, and of what kinds.

Now the episode is not to everyone's taste. Indeed very little criticism has been written about it, relatively speaking, and much of it, especially early on, was hostile. Most felt bored reading prose that seemed to be trying to be boring. Those who took the trouble to write about it at all spoke vaguely of "stock-phrases," "commonplaces" and "fossilized expressions," and latterly have tried to allegorize the episode as an attack on the post-modernist subject, and the constraints of discourse in a colonial city and a capitalist economy. Several wrote on the language itself, but they tended simply to identify representative patterns within Bloom's speech, including a propensity to cliché. And they proceeded selectively, usually noticing only the clearly opaque or figurative idioms, that is, the most obvious ones. None set themselves the task of identifying all 100% of the clichés, though one critic estimated there were probably at least 700 (Raleigh, 1981).

So that's where the database I've created comes in. It occurred to me that this must be the most difficult of Joyce's episodes to translate, first because of the apparently high number of clichés, and second because many of them would be completely opaque. As we'll see, it may in fact be the most nightmarish episode in the book for translators to deal with, though for different reasons than we had anticipated. And of course *Ulysses* is now being translated, or translated *again*, into every major world language. I decided to help the translators by identifying all the clichés and putting a searchable database up on the web. As a native speaker of English, I had no doubt I'd be able to identify virtually every cliché, and I used Google and a number of dictionaries of idiom and cliché for anything I wasn't sure about. The initial goal was simply to ID all the clichés, however many or few, though by this time I'd had an intimation there were rather more than we had suspected.

2. Phraseology

But what's a cliché? It's fortunate for this project that a sub-field of linguistics has grown up in the last 25 years that can answer this question. The sub-field is phraseology, and though of course pre-alphabetic people were quite as aware of formula phrases as any phraseologist, we may say that phraseology was "invented" only in the late 20th century. Anthony Cowie (1994) defines it as «the study of the structure, meaning and use of word combinations», and it's had two branches virtually from the beginning. Russians, Middle Europeans and French

tend to be interested in fixed forms like proverbs and idioms, and even more so in what they consider *opaque* idioms, the ones in which the words don't add up to the meaning of the whole expression (Granger and Paquot, 2008), as with "cooked his goose" or "smell a rat". Such combinations are called "non-compositional."

The British and American branch of phraseology interests itself more in the looser, more transparent expressions like collocations, in which the words retain the meanings they have when standing alone, as with "criminal propensities" and "highly advisable." Such combinations are called "compositional". Most combinations fit somewhere along a continuum, with fixed, opaque forms at one end and the loosest, most transparent collocations at the other (Granger and Paquot, 2008).

The continental researchers tend to define word combinations grammatically, then look for them, but the Anglo-Saxon investigators, armed with computers, simply scan through digitized corpora for high-frequency "co-occurrences," irregardless of linguistic character. Such research demonstrates that the looser forms, especially collocations, are omnipresent in normal discourse, while the traditional fixed and opaque ones are fairly rare. If you were raised on transformational-generative grammar you may have come to think of language as modular – composed of a grammar and a lexicon, with the grammar lining up individual words syntagmatically, and the lexicon dropping paradigmatic vocabulary items into these syntagmatic slots.

But a phraseologist like John Sinclair sees the grammar managing not just words but multi-word lexical units as well – strings of collocations, idioms and other formulae, each of which is chosen as if it were a single item (Sinclair, 1991). This is pretty much what Leopold Bloom's monologue looks like, as he tries to impress Stephen, the University College B. A., with his Polonius-like sententiousness. Or rather, it's the basis of Joyce's style in this episode, as if he were a phraseologist *avant la lettre*.

3. Identifying Clichés/Phrasemes

As Alison Wray (2002: 19) points out, «the relationship between definition and identification is circular» in phraseology. We need good examples before we can elaborate a definition, and we need a definition before we can find good examples. Distributionists use computer searches to find co-occurences, but «it is often intuitively clear that some patterns are more important and relevant than others». Wray attempts to resolve this conundrum by defining phrasemes (or "formulaic sequences") as those "prefabricated strings" stored holistically, as a unit, rather than being generated analytically from words and morphemes with grammatical rules.

I myself have used both sides of this circular process. I identified every phraseme I could in "Eumaeus" simply by native-speaker intuition, but then checked my intuition several ways. First, I searched the phraseme in Google, and if I found a few hundred thousand examples, I concluded I was looking at an incontrovertible instance. If I found five thousand examples, I concluded I was looking at a more rare, but equally clear instance. Some phrasemes produced few enough hits that I resorted to various dictionaries of cliché and slang, or inspected regular dictionaries for their appearance in examples under the entry for a component word. I've left out a number of phrasemes I suspect were Victorian-era clichés that have passed into oblivion now, except among literature majors reading in the period, and included a few that I was "intuitively" sure of, as a native speaker, even though Google found only a handful of instances. In fact, Google agreed with virtually all my intuitions, and no more than a handful of phrasemes are present in the database almost entirely on the basis of my intuition.

4. Developing a Taxonomy of Lexical Types

There is a superfluity of taxonomies of lexical types. Indeed there are so many taxonomies that there's a great deal of confusion and miscommunication in phraseology as a whole, for lack of a common terminology (Wray, 2002: 45). Wray adduces a partial list of 21 of the best known, and divides them loosely into ones for either theoretical or practical projects. A taxonomy that is theoretically driven, she says, «needs to justify itself with reference to an external model of what language is and how it works. In most cases, this is an assumption that the most creative language is assembled by rule from small components, so that there is a relationship between formulaicity and an absence of creativity». But practically-driven taxonomies «do not need to be theoretically grounded or theoretically robust. They do, however, have to work for their intended purpose» (Wray, 2002). She has in mind here both dictionary making and language teaching. Language teachers, for example, need to present phrasemes in an orderly manner, by type, for pedagogical purposes, and student success justifies the taxonomy, no matter how it was confected. I would add that translation is another such practically-driven project, and this database will justify itself, first, if translators find it useful, and second, if it helps us to describe more of Joyce's art in "Eumaeus" in our critical interpretations.

Sylviane Granger and Magali Paquot (2008), in an attempt to reconcile traditional and distributionist terminology, have elaborated H. Burger's 1998 taxonomy of "phraseological units" into a more comprehensive one of their own. I in turn have adopted and slightly adapted this "big-tent" classification because it commodiously describes the variety of lexical types in "Eumaeus" and this project is of course text-driven and practical rather than theoretical. I divide her "idiom" category morphologically into Complex, Compound and Single units, partly for more granularity, and partly because Joyce uses 108 Single-Unit clichés, entities that would ordinarily escape a phraseological analysis. I've added a "Titles" category, for the names of poems, songs and novels that have entered common parlance to the point of cliché.

5. Description of the Database

I created a separate record in Filemaker for each of the clichés I identified, and added descriptive fields, the first of which of course is Lexical Type. To accommodate a translator's further requirements, I've added fields that tag each cliché in terms of its Semantic and Syntactic types, Style Level, Domain, and Language (Joyce uses clichés from several languages).

The Lexical Type field identifies Collocations, Adverbial units, and Single, Compound and Complex units, with additional categories for less common occurrences like Sentence Stems, Routine Formulas, Lexo-Grammatical Collocations, Sentence Idioms, Speech Acts, Quotations, and of course Titles (see Tab. 1, for examples, of each type).

The Semantic Type reports a crucial attribute for translators, telling them whether they're looking at phrasemes that are opaque, like "a grass widow", figurative, like "drew the line," or transparent, like "nobody to blame but themselves". The Syntactic Type indicates whether the expression is fixed, like "there was nothing for it" or "not to put too fine a point on it" or has variants, like "in view of the hour/time" or "once in a while/way" or is simply a free combination (collocation) of words that attained the status of cliché by their frequent co-occurrence, like "very badly needed" and "a bit unsteady".

The critical Style Level field indicates the cliché's register, along a spectrum from "formal" through "ceremonious", "neutral", "informal", "slang" and "vulgar." A Domain field indicates that the expression is part of a terminology, whether medical, nautical, governmental, etc., and

the Language field indicates the variety of English being analyzed (Irish, American, British, Archaic British), or the language of foreign clichés in the episode (French, Italian, Russian, Greek, etc.).

6. Results

This phraseological database establishes, first, that there are far more clichés in "Eumaeus" than anyone has heretofore ventured to suggest. As it turns out, there are at least 2.561, perhaps 2.000 more than most of us suspected. In the edition I was using, the maximum number of clichés on a single page was 53, with an average of 36. This leaves, in fact, hardly any room on the page for phrasing that *isn't* clichéd. "Eumaeus" is in effect a thesaurus of cliché, an encyclopedia of Dublin common parlance, every sentence stuffed to bursting, like a paté goose, with the habitual locutions of a somnolent imperial province. Joyce has been systematic, and comprehensive too, with virtually every lexical type you'd care to name being represented (Tab. 1). From a linguistic perspective, "Eumaeus" is a phraseological plenum of Dublin English circa 1904.

Lexical Type	Num	%	Examples
Collocations	1083	42	a necessary evil, a blithering idiot
Adverbial Units	533	21	on the prowl, for every contingency
Complex Units	251	10	made tracks, swear a hole through a ten gallon pot
Compound Units	166	6.5	a golden rule, heir apparent, dark horse
Single Units	108	4	the needful, bloody
Sentence Stems	156	6	as it so happened, he meant to say, God knows
Routine Formulas	150	6	if I were in your shoes, Is that so?, It beats me.
Lexo-Grams	46	2	blessed with, tantamount to
Sentence Idioms	13	.5	ignorance is bliss, let bygones be bygones
Speech Acts	2	.078	on my solemn oath, I grant you
Quotations	41	1.6	our daily bread, at the eleventh hour
Titles	12	.46	the last of the Mohicans, our mutual friend

Table 1: Lexical Types

The second major finding is that only 151 expressions, or about 5.9% of the database, are opaque (Tab. 2).

Semantic Type	Num	%	Examples
Opaque	151	5.9	making both ends meet, to break the ice
Figurative	335	13.1	out of his depth, rule the waves, a roaring trade
Transparent	2074	80.9	I beg to differ, solid food, a pardonable weakness

Table 2: Semantic Types

But even fewer are also fixed in form, their fixedness making them even easier to recognize. Of the fixed, opaque expressions, 35 are complex units like "horse of another colour" and "make ducks and drakes of", 38 are compound units like "knockingshop" and "weather eye", and 26 are adverbial units like "off the reel" and "all at sea". So a mere 99 or 3.8% of Eumaean clichés are the "pure" sort, that is, not only opaque but fixed in form, and altogether easy to recognize. There are 335 figurative phrasemes in the database as well, like "sailing under false colors" and "the cut of his jib" and translators usually spot them without trouble.

But the huge preponderance of cliché in "Eumaeus" is in fact transparent, that is, invisible as cliché, often even to those born into the language. When I ask a classroom full of undergraduate English majors to underline the clichés on a page of "Eumaeus", they'll usually spot from 4 to 6 to 8, ignoring anything that doesn't have some rhetorical effect to distinguish it, like alliteration or chiasmus or antithesis, or some gnomic aura of ancient wisdom. If native speakers have this much trouble spotting transparent phrasemes, we can predict a certain difficulty for most translators, however steeped they may be in English as a second (or third) language. A translator might look at "the power of pelf" and suspect, from the alliteration, that it was a Victorian commonplace. Indeed it was, and it's still used today in India and Pakistan, where it has become a journalistic banality. But what about "inherent delicacy"? Only a well-read native speaker is likely to recognize it as a rare and elegant collocation. It isn't lexicalized (doesn't appear in dictionaries) so a translator can't look it up. He or she may well render it verbatim, into a phrase that isn't clichéd in the target language, much to the impoverishment of the translation. "Eumaeus" is a summa of cliché, but not of the "pure", fixed, opaque kind that so insistently solicit the translator's eye. Joyce's clichés are the phrasemes of everyday speech, the ones that people use without consciousness, without trying to express themselves explicitly in traditional stock phrases. It's a language of free, non-compositional transparencies – collocations, adverbial phrases, routine formulas and sentence stems that don't wear their banality on their sleeves. Fully 81% of Eumaean phrasemes are free, non-compositional and disarmingly transparent. For translators, "Eumaeus" has to be the most ferocious obstacle course ever written.

Finally, it's crucial that a translator know the cliché's style level, sometimes called "register." Is the cliché formal? ceremonious? neutral? familiar? vulgar? Is it slang? To mistake a cliché's style level would drastically change the tone of a sentence, and to get a number of them wrong would make hash of the social "character" of the character speaking them. Joyce's *Ulysses*, as much as Jane Austen's *Pride and Prejudice*, is a "novel of manners", which is to say a novel about the multitude of linguistic registers, dialects, idiolects, polite formulas, punctilios, professional locutions, plain speaking, euphemisms, evasions, indirections, vulgarities, familiarities and insinuating usages that create the social drama. To get these nuances wrong would be to miss the "manners" themselves.

I noted before that many critics had found "Eumaeus" boring. But a number have always found it hilarious, citing for example the "incongruities of register" (O'Neil, 1996) that seemed comical. Just how pervasive these must be, we see in Tab. 3.

Style Level	Num	%	Examples
Formal	86	3.3	the duty devolved upon him, suitable ways and means
Ceremonious	443	17.2	women of ill fame, cordially disliked
Neutral	1478	57.7	so far as he could see, pressed for time
Informal	510	19.9	a stonesthrow away, on the spree
Slang	39	1.5	sheer cussedness, greenhorns, bunged up
Vulgar	5	0.1	bloody, I don't give a shite

Table 3: Style Level

Bloom operates from a broad base of 1478 tonally neutral clichés, then works up into the ceremonious and formal zones 529 times, and down into informality, slang and vulgarity a further 554 times. The upper and lower zones of his range are almost exactly symmetrical.

So what is the "social scene" these numbers underscore? A commonplace man of little schooling is trying to impress University College's most brilliant new B. A. (who has a degree in modern

languages). He pours forth a flood of phrasemes he's apparently stocked up on over the course of an un-heroic and uncultured lifetime. In effect, a social nonentity with a phraseological repertoire largely in the "neutral" zone of the language's most colorless clichés (about 60%), aspires above his station to eloquence, to the ceremonious and formal register (about 20%), and crashes down from there repeatedly into the toneless middle zone, or further down, into domestic informality, slang, and real vulgarity (about 20%). Each tonal ascent betrays him as a *poseur*, a bounder, a petit bourgeois with delusions of learning, refinement and finesse, and each bathetic collapse into informality confounds his presumption and unmasks him utterly. Joyce's art inheres entirely in this drama of failed impression management. Bloom struggles for dignity, authority, and comradely solidarity with Stephen. His failures are bathetic when we laugh at him, and pathetic when we feel for him. Bloom is both burlesqued and endeared to the reader in his desperate and bravura performance. Joyce makes high drama out of this compendium of the ordinary, and creates a complex social portrait in the midst of a hail of commonplace.

7. Conclusion

This database alerts translators first to the great and largely unsuspected number of clichés in "Eumaeus." More importantly, it alerts them to the fact that fully 81% of the phrasemes are transparent ones, mostly collocations, that can't be correctly translated verbatim. Rather, they need to be translated carefully into equivalent clichés in the target language, lest the concentration of cliché be diluted. Secondly, it identifies the material basis of Joyce's art, which proceeds at the micro level, as one stock phrase collides tonally with the next, producing a comedy of manners. It shows the translator exactly what's happening, phraseme by phraseme, as *Ulysses'* humble hero tries desperately and vainly to speak above his natural linguistic register. Finally, it provides a statistical x-ray of Joyce's art, enabling critical observations that would be impossible without the database.

References

Cowie A. P. (1994). Phraseology. In Asher, R.E., editor, *The Encyclopedia of Language and Linguistics*, Oxford: Oxford University Press, pp. 3168-3171.

Granger S. and Paquot M. (2008). Disentangling the phraseological web. In Granger, S. and Meunier, F., editors, *Phraseology: An Interdisciplinary Perspective*, Amsterdam-Philadelphia: John Benjamins.

O'Neil C. (1996). *Too Fine a Point: A Stylistic Analysis of the Eumaeus Episode in James Joyce's* Ulysses. Trier: Wissenschaftlicher Verlag.

Raleigh J.H. (1981). On the Way Home to Ithaca: The Functions of the 'Eumaeus' Section in *Ulysses*. In Bowen, Z., editor, *Irish Renaissance Annual*, London-Toronto: Associated University Presses.

Sinclair J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Wray A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.