Dylan Glynn

The conceptual profile of the lexeme *home*: A multifactorial diachronic analysis

**Abstract:** Despite the descriptive power of the Idealised Cognitive Model (Lakoff 1987), the analytical framework faces two inherent problems. First, Idealised Cognitive Models treat language-culture to be a homogenous object study, producing ‘idealised’ results that do not readily account for social variation or change. Second, they produce results that are not systematically falsifiable, an essential tenet of scientific method. In a diachronic study of the American concept of *home*, this study seeks to develop analytical tools for the empirical description of conceptualisation that produces results sensitive to social variation and that can be falsified. Employing the profile-based usage-feature method (Geeraerts et al. 1994, Gries 2003), the study examines samples taken from texts by three 19C American writers (James Cooper, David Thoreau and Fredrick Turner) and two 20C lyricists (Woody Guthrie and Bruce Springsteen). The aim is to determine whether usage-feature analysis is capable of capturing the kind of conceptual structure typical of studies on Idealised Cognitive Models. The analysis focuses on a set of metaphoric source concepts and a range of usage-features chosen as indices of conceptual structure. Using multivariate statistics, it investigates the relationship between the different metaphors relative to their usage over the course of two centuries. The study demonstrates the proof-of-principle that socially sensitive and falsifiable descriptive studies of culturally determined conceptual structure are possible.

1 Introduction

The Idealised Cognitive Model, proposed by Lakoff (1987), represents a powerful descriptive tool for making generalisations about language, society and cultural worldview. However, the analytical framework it employs and the methodological premises it assumes face two fundamental limitations. Firstly, the ‘idealised’ nature of the proposed language-culture structures is at odds with the theory’s usage-based assumptions (Langacker 1987) and does not produce descriptions that are sensitive to social variation. Secondly, the analytical
method of lexical semantic co-occurrence lacks a means for the falsification of the findings it produces, a limitation that runs contrary to the broader methodological assumptions of empirical science. Despite this, the study does not argue that the analytical method of Idealised Cognitive Models is inherently flawed. Indeed, its role is essential in developing hypotheses about socio-cultural-linguistic structure. Instead, it is argued that we need to develop an empirical methodology designed to test hypotheses proposed with the Idealised Cognitive Model framework.

The goal of this study is to demonstrate the feasibility of using multivariate usage-feature analysis for the description of conceptual structures in language and culture. This method was developed by Dirven et al. (1982), Rudzka-Ostyn (1989), and Geeraerts (1990). However, the application of multivariate statistics to the results of the analysis is the step that gives the method its descriptive power. Drawing on established quantitative methods in sociolinguistics, Geeraerts et al. (1994, 1999) and Gries (1999, 2003) developed the use of such categorical multivariate techniques. In recent years, the method has gained popularity and is termed the profile-based approach by Gries and Divjak (2009), Divjak (2010a, 2010b) and Deshors and Gries (2014) and multivariate usage-feature analysis by Glynn (2009, 2010a, 2010b, 2014c, 2014d, submitted), Krawczak and Kokorniak (2012), Krawczak (2014a, 2014b), Fabiszak et al. (2014) and Klavan (2014).

The concept of home is a fundamental one in Germanic languages and culture. This study draws on both qualitative and quantitative methods in an attempt at describing the concept and its diachronic variation in 19th and 20th century Anglo-Saxon American culture. It bases its analysis on two specific genres and a single lexeme, but is part of a larger project that examines a range of lexemes and genres (Glynn to appear). The results presented here include occurrences of the lexeme home in J. Cooper (1789–1851), H. Thoreau (1817–1862), and F. Turner (1861–1932), for the 19th and from the ballads of W. Guthrie (1912–1967) and B. Springsteen (1949–) for the 20th century. Although these sources and the text types are distinct, their place in two diachronically distinct contexts is arguably comparable.

Limiting the study in such a way makes it possible to better control for stylistic effects, but limits the representativeness of the findings. It is argued that by selecting the sources in this manner, we can be precisely sure of what our results represent, even if this is at the cost of being able to make broad generalisations. In order to establish how general the results are, future studies, taking divergent sources, would need to be undertaken.

Section 2 presents the data and the analysis. The results of the analysis and their interpretation are found in section 3. The discussion ends with a summary in section 4.
2 Data and Analysis

2.1 Lexeme and concept

Both the concept home and lexeme home (hjam Northumbrian; hjem Cumbrian; haim Scots, Heim German, heim Icelandic, hem Swedish, Danish) hold a special place in Anglo-Saxon culture and, indeed, Germanic culture and languages more generally. The Proto Germanic origin *haimaz, like its contemporary counter part, appears to be an abstract concept, not restricted to a building and ultimately derives from the Proto Indo-European root *tkei- ‘To lie, settle down’. The abstract nature of the meaning of the lexeme is evident synchronically in how it is extended to include a country, a street, a village or, in fact, any place where one holds an emotional attachment, perhaps related to one’s origins or a feeling of personal security. The same is true diachronically and this is evident in the wide range of uses and variants, across the Germanic languages, which are not restricted to a specific lodging or building.

In this, the lexeme is distinct from that of house (Haus German, huis Dutch, hus Swedish and Danish, hús Frisian and Icelandic, house/hoose Scots, *husan Proto Germanic), whose ultimate origin is uncertain, but which profiles the function of a building, its role in sheltering and protecting. This understanding of the word is in line with a proposal expressed by various dictionaries that the word derives from a proto Indo-European root *keudh- ‘hide’. (Cf. West Germanic *hudjan). Importantly, the lexical semantic distinction between house and home is present in all but one Germanic language, while it is effectively absent in other European language families.¹ The casa and dom roots in the Romance and Slavic languages typically mean ‘house’ or even ‘building’, but not specifically ‘home’.²

¹ Modern Dutch and Frisian have no cognate, but have modified a form of the cognate for house to indicate ‘home’. The lexemes huis and hús ‘house’ have produced thuis and thús respectively and mean something closer to the abstract concept of home than ‘house’ per se. The existence of the variant is, perhaps, testimony to the importance of the concept. However, of course, that the original cognates for home, heem (Middle Dutch) and hem (Old Frisian), were lost could be offered as a counter argument. Note also, that the South Slavic languages, Bulgarian, Macedonian, Serbo-Croatian and Slovenian, have the distinction between ‘home’ and ‘house’, drawing on the Germanic root in the alternation. For example, in Slovenian, the lexeme doma refers to the English equivalent home, while the lexeme hiša indicates ‘house’. It is important to note that these observations should not be taken to suggest that other European languages lack a means for expressing the concept of home. Often, a partially lexicalised phrase fills the role, such as u sebja in Russian or chez soi in French, both of which express a concept of being at one’s own place, wherever that may be.

² Cf. casa Spanish, Italian and Catalan, casă Romanian, chambre French, domo Sardinian, dom Polish and Russian, dim Ukrainian, and dům in Czech.
However, the aim of this study is to go beyond lexical semantics. It seeks to operationalise the Idealised Cognitive Model in order to render falsifiable the broad cultural generalisations typical of the work of Wierzbicka and Lakoff. A cloudless flight across Europe will reveal the transition from the sparsely placed ‘hams’ of northwest Europe to the village clusters of southwest Europe. In the north, large houses surrounded by fields, fenced off from their neighbours by a close row of trees are in direct contrast to the tightly knit clusters of the idyllic Mediterranean villages. Although such cultural generalisations are possible, when based on ad hoc observations of lexical semantics or even town planning, regardless of how true or informative, such generalisations do not make social science. Instead of examining lexical semantics per se, we examine, systematically, the contextualised use of the lexemes, in the hope that generalisations about culture can be made in way that is not only falsifiable, but also sensitive to diachronic social variation.

### 2.2 Choice of data

This study seeks to develop methods for the description of cognitive models. The diachronic dimension that this paper focuses upon is an example of the kind of descriptive challenge an empirical method for conceptual analysis must be able to meet. In total, 300 occurrences, 150 from both the 19th and 20th centuries, were taken with substantial context. The relatively small number is due to the fact that examples must be manually analysed. Moreover, the choice of sources for the data is somewhat unusual and warrants justification.

Rather than examine a diachronic corpus, such as Davies (2010), five specific writers were chosen. There are two justifications for this. Firstly, the inherent limitation of corpus-driven research is that every occurrence is treated equally. In other words, each use of a form is assumed to have the same value or degree of contribution to the language-culture system. Although a valid operationalisation, such an assumption cannot completely explain language, cultural knowledge or conceptual structure. Certain occurrences, due to differences in perceptual salience or cultural relevance, carry more or less weight. By biasing the data in favour of certain types of examples of language use, the study is an experiment that seeks to respond to this methodological limitation.

---

3 The data were extracted and cleaned by Joakim Sten. An entirely distinct analysis of the data is presented in Sten and Glynn (2011).
Secondly, having explained the use of specific authors, we need to consider why we chose authors from different genres. When performing diachronic studies, one normally seeks to obtain data from a single text type or genre. Despite the sound reasoning of this line of methodology, it entails an inherent circularity. Firstly, a genre is a part of the time it comes from. Although some genres do persist over long periods, they all eventually die or become something entirely different. The novel, in the contemporary understanding of the term, is a relatively new genre, just as news press. Already, many would consider, rightly or wrongly, the novel to be on track to be replaced by other media. News press is also changing and in such a dramatic way that it is becoming difficult to speak of a single genre diachronically. The broadsheet is already dead in the English speaking world and the six o'clock news, online news sites and editorial blogs are in line to end Fleet Street, even within the current generation. Those few genres that have persisted over longer periods of time, such as poetry and theatre, are arguably archaic if one is seeking linguistic evidence for a contemporary conceptualisation of the world. That is not to say that they are uninformative, but there is an argument to be made for taking a genre that is typical of its time, rather than a genre that exits in two different times. In other words, by keeping genre a constant in a diachronic study on culture or conceptual structure, one may actually weaken the representativeness of the data. For this reason, texts that are representative of the distinct periods are chosen rather than selecting texts that maintain genre consistency. Due to this choice, care must be taken not to interpret stylistic variation in terms of any underlying cultural or conceptual structure.

For the 19th century, the works of two writers are taken: James Fenimore Cooper, a popular writer of historical romances about the nation building era of the United States of America and Henry David Thoreau, a popular philosopher who was also concerned with the topic. A single text, *The Frontier in American History* (1893), by Frederick Jackson Turner, written somewhat later, though still within the culture and times of the 19th century, is also included. It is felt that this text is in keeping with the style and era concerned and addresses especially the concept of home in the nation building of North America.

Turning to the 20th century, the idea is to work with texts that have a comparable socio-cultural role to the texts of the 19th century. To these ends, we draw upon the popular ballads of two writers who are held to be spokespeople of their generation and nation. Like their 19th century compatriots, the popular songwriters Woodrow Wilson Guthrie and Bruce Frederick Springsteen are concerned with their nation, its wellbeing and identity (Shelton 1986; Marsh 1987; Cray 2004). Moreover, the concept of home, both in the personal sense and in
the society sense of homeland, is an important theme in their work (Partridge 2002; Guterman 2005; Cowie and Lauren 2006; Jackson 2007; and Lifshey 2009). Songs such as “This land is your land” (Guthrie 1940), “Pastures of plenty” (Guthrie 1941), “Bound for Glory” (Guthrie 1942), “Hard travelin’” (Guthrie 1944), or “Born to run” (Springsteen 1974), “Backstreets” (Springsteen 1975), “The river” (1979) and “Born in the USA” (Springsteen 1982) would be known, as least passively, to most contemporary English speakers. Like so much of their work, these examples treat the notion of home, to varying degrees, in relation to the individual, to their generation and even to the nation.4

2.3 Usage-Feature Analysis

The aim of the feature analysis is to operationalise semantic profiling in such a way that the use of the lexeme (as opposed the the lexeme itself) can be treated as an index of conceptual structure. Before we exemplify the actual usage-feature analysis, we begin with some examples of the texts in question. Examples (1) to (5) are chosen to represent the diverse styles, yet similarity in theme across the authors and periods. It should be noted how the examples include instances of use that go beyond the strict lexical semantics of home. It is an essential part of the proposed methodology that the emerging picture is one of the cognitive model in the broad cultural sense and not of the concept associated with the lexeme in any strict sense. The italics are added.

(1) Others left the country; seeking in that place they emphatically called home, an asylum, as they fondly hoped, for a season only, against the confusion and dangers of war.
(Cooper, The Spy: A tale of the neutral ground)

(2) But the place which you have selected for your camp, though never so rough and grim, begins at once to have its attractions, and becomes a very center of civilization to you: Home is home, be it never so homely.
(Thoreau, Canoeing in the Wilderness)

(3) The Mississippi Valley has been the especial home of democracy. But the democracy born of free land, strong in selfishness and individualism.
(Turner, Frontier in American History)

---

4 For sake of consistency, Guthrie’s novel House of Earth is not included in the sample.
(4) Yes, we ramble and we roam. And the highway that’s our home. It’s a never-ending highway. For a dust bowl refugee.
   (Guthrie, *Dust Bowl Refugee*)

(5) Sent me off to a foreign land. To go and kill the yellow man. I come back home to the refinery. Hiring man says “Son, if it was up to me”.
   (Springsteen, *Born in the USA*)

The challenge for a quantitative analysis is to operationalise the kind of meaning carried by this lexeme in such a way that falsifiable generalisations can be made. In order to approach this question, we need to think about how home is being used to profile different dimensions of the concept home. In example (1), the referent of home is a ‘country’, that is a political entity associated with a physical place. Whether this should be treated as a metaphor, a metonym, or a literal usage depends on how one defines the lexeme home in the first place. If the literal understanding of home is a ‘house’ where you feel safe or where you grew up, then the reading is metaphoric. If the literal definition of home is a place where you feel safe or from where you originate, then the reading is metonymic. If home is defined as any association of belongingness or safety, then the reading is literal. We will leave this question aside for the moment (q.v. section 2.3.1).

Thoreau’s example of home (2), typical of the complex relationship between ‘land’ and ‘home’, speaks of taking untamed land, which is not a home and making it one by taming it. This example not only highlights important characteristics of what home is (and is not), it appears to speak of one’s natural desire to make any place safe and homely, regardless of where it is or how unfriendly the context. The third example is from Turner who is later than Thoreau or Cooper and stylistically quite distinct. As a historian, his discourse is much more concerned with politics and law, but his examples are narrative descriptions of the trials and tribulations of home and land. In example (3), land is the source concept for home, regardless of what kind of conceptual relations that entails, but it not the home of a human, instead of an abstract idea.

The songwriters of the 20th century are, of course, markedly different in style, but they share the same concern for the concept at hand. In contrast to the role of ‘land’ in the 19th century, the relationship between streets and home is especially important in the 20th century texts. At a denotation level, it is far from obvious why streets would be chosen as a counter part to the 19th century land. However, in the urban 20th century, when we are discussing a social space, that can be your home by heritage, or can be made your home, a place
you belong, the concept of STREET, in its metaphoric or metonymic usage, is a comparable concept.

Example (4) serves as a good example of how the concept of STREET, here profiled as highway, is a source domain for HOME. Such questions are not treated in this study, but are important in the broader research project (cf. Glynn to appear). In Springsteen’s example (5), home is used in an abstract sense, but not linked to LAND, NATION or HOUSE, but an ABSTRACT PLACE, here metonymically represented by an oil refinery. The individual comes home from war, to find he has no home because his job has gone. The lexeme home is understood here in a complex metaphoric and metonymic sense. The refinery is, in fact, not the home, but it represents security and origins, standing for HOME metonymically.

2.3.1 Conceptual structures of similarity and contiguity

The feature analysis of the conceptual structure faces several challenges. Instead of identifying metaphors and metonymies, the analysis simply identifies what are termed ‘source concepts’. This is in order to avoid the issue of distinguishing conceptual structures of similarity and contiguity, which proved impossible to adequately operationalise. Let us briefly consider why.

Firstly, the concept of HOME was often found to have a complex relationship with the concept of LAND. Consider examples (6a) and (6b).

(6) a. America was not simply a new home; it was a land of opportunity.
   (Turner, The Frontier in American History)
   b. ... rapid conquest of the wilderness. We have so far won our national home.
   (Turner, The Frontier In American History)

The complexity arises from the fact that the notion of ‘home’ itself is so abstract that almost anything can be used to indicate home, or at least a feeling of home. Example (7) demonstrates why this is important.

(7) Now honey, I don’t wanna clip your wings. But a time comes when two people should think of these things. Having a home and a family. Facing up to their responsibilities.
   (Springsteen, I Wanna Marry You)

Are these examples figurative? At first, we might assume them to be literal. However, in the context of the genres in question, it is a reasonable argument
that the author is expecting the audience to extrapolate from this literal reference in a metonymic, and perhaps even, metaphoric manner. In example (7) the meaning of *home* is the physical place where one feels safe and experiences a sense of belonging. At some level, of course, the relationship between these two facets of the meaning is figurative. In a given example, if one is foregrounding the actual place and backgrounding the abstract associations that determine the place, then the reading is metonymic – the place stands for the association. A quotidian expression such as *I’m going home* would be an example of this kind. In such a situation, the place one feels safe is foregrounded and the state of feeling safe is backgrounded, the place metonymically standing for the abstract concept. However, if one is foregrounding the abstract concept and the concrete place is not the referent, then we can interpret this as a metaphor. For example, the idiom *wherever I leave my hat is home* is clearly metaphoric, even if the source domain could be interpreted as metonymic; the place that you ‘leave your hat’ standing for the place you ‘feel at home’. The complexity arises in natural examples, such as (7). Springsteen is speaking in abstract terms, with no specific physical shelter in mind. This is made clear by the reference to ‘clipping of wings’, where freedom is set up as the opposite of the staid concept of HOME and reiterated by the collocation – *having a home*, which, like *having a family*, is often associated with metaphoric uses. This example, which is both relatively straightforward and typical of the usages of the lexemes in the sample, poses a serious analytical question. Are such examples literal, metaphoric or metonymic? If they are figurative, what then are the source domains or broader concepts that are activated?

The metaphoric structuring in examples (8a) and (8b) should be clearer.

(8) a. Tonight I got dirt on my hands but I’m building me a new home.
   (Springsteen, *Lucky Town*)

b. It meant to them, as to the American pioneer that preceded them, the opportunity to destroy the bonds of social caste that bound them in their older home, to hew out for themselves in a new country a destiny proportioned to the powers that God had given them, a chance to place their families under better conditions and to win a larger life than the life that they had left behind.
   (Turner, *The Frontier in American History*)

In example (8a), the metaphor is that of hard work being the foundation for a HOME. Such examples of the Protestant work ethic abound in the sample but it is not always clear whether this is metonymic or metaphoric. Example (8a) is of the same kind as that identified by Goossens (1990), which he terms met-
apthonymy. Since it is not always clear whether the source concept is a part of the target concept or distinct from the target concept; whether the conceptual relation is one of contiguity or similarity is difficult to determine. In natural dialogue, unless the speaker intends a pun or blend, such ambiguity is probably quite rare. However, in literary texts, such as those we are considering here, such ambiguity is commonplace. In (8a), the agent is literally working and literally getting his or her hands dirty doing so, but we cannot know whether this is literally part of building a shelter that will serve as a home or whether the physical labour is distinct from the abstract nature of home building, but metaphorically designates it, by, for example, going out to work every day to earn money to buy a home. Indeed, in (8a), it is likely that dirt is used metaphorically for criminal activity, adding further to the complexity of the interpretation. Knowing that this is literary text and thus necessarily decontextualised, such examples are inherently both metaphoric and metonymic and are probably intended as such by their authors.

It is not the point of this study to delve into the intricacies of metaphor analysis. Despite the importance of such a discussion, our concern here is a simple operationalisation of these subjective notions that will enable repeatable and falsifiable results. For these reasons, no effort is made to distinguish metaphor and metonymy in the analysis. For each example, the target concept, be that an independent concept (metaphor) or a dependent ‘part’ of the source concept (metonymy), is annotated. The same principle is applied to the source concept. When the lexeme \textit{home} refers to a ‘house’, the source concept is treated as \texttt{house}, when it refers to ‘land’, it is treated as \texttt{land} and when it refers to an abstract place, it is annotated as \texttt{place}. By pairing the two at the end of the analysis, we know what two concepts are involved in each example, but not whether their relation is one of similarity of contiguity. Further details on exactly what concepts are identified are offered below in section 3.1.

Example (8b) represents yet another problem in the conceptual analysis. Here we have two conceptual profilings of \texttt{home}. The first, similar to that of (7), is of a constrained existence: the experiencer is ‘bound’ to their ‘home’ by ‘social caste’. However, in the example, a second conceptualisation of \texttt{home},

\begin{table}
\centering
\begin{tabular}{lccccc}
\hline
 & place & house & land & nation & Total \\
\hline
19C & 42 & 57 & 29 & 11 & 139 \\
20C & 61 & 61 & 14 & 12 & 148 \\
Total & 103 & 118 & 43 & 23 & 287 \\
\hline
\end{tabular}
\caption{Frequency of source concepts relative to period.}
\end{table}
contrasted with the previous, is also presented. The Protestant belief that hard work is the basis of a good life is overtly indicated with the reference to God and the metaphor of hewing out a nation, which we can read as 'building a home'. So, which metaphor is relevant here, the work-build metaphor or the constraint-limit metaphor? In such situations, the metaphor directly associated with the lexeme in question in analysed and not any others. This leads to situations where the more conceptually rich metaphor is omitted from the study, but following this principle systematically is the only way of operationalising the analysis.

The systematic analysis of all the examples revealed the following source concepts: House, Land, Nation, Person, Street and abstract place. The concepts Person and street were only found in the 20th century and were infrequent. These examples are not included in the study. The frequencies of the remaining examples, distributed across the different source concepts, are presented in Table 1.

2.3.2 Semantic usage-features

The theory of Idealised Cognitive Models is not exclusively about metaphoric and metonymic structuring. Indeed, on the contrary, it is an abstraction across representations of the world, indexed by language structure, and, from a usage-based approach, across the usage of language. Usage-based cognitive models, based on re-occurring instances of actual language use, permit the addition of formal and semantic characteristics to the description of metaphors: just as each utterance is analysed for its source and target concept, a range of other usage features are annotated.

For example, two formal and objectifiable identifiable categories, the main verb and the preposition of the home noun phrase, can be used as clues to the semantic nature of the source domain. In the above examples, the prepositional and verbal collocates, 'have a home', "in the land", "build a home", "in their home", when identified across all the occurrences can be found to be associated with or indicative of certain metaphors and or metonyms. This approach is not employed here, but is used in Glynn (to appear). In the current study, we focus on a range of purely semantic features.

The semantic analysis is based on determining the designatum for each occurrence and then through a semantic analysis of the context, a set of semantic features are ascribed to the utterance. The semantic features are what the Russian tradition would term conceptual analysis (Stepanov 1997 and Vorkachev 2007 inter alia). In the Cognitive Linguistics tradition, they are simply
encyclopaedic semantic usage-features. Systematically identifying such features is the basis of the methodology and was established by Dirven et al. (1982), Rudzka-Ostyn (1989) and Geeraerts (1990).

Each of these semantic usage-features is operationalised with a simple question. There are three possible values in response: (i) the attribute in question is profiled in the example; (ii) the absence of the attribute is profiled in the example, (iii) the attribute is not profiled (its absence or presence) in the example. This third category is important since, in many examples, not all the features are applicable. The semantic features include:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lodging</td>
<td>Is the designatum serving as a place to ‘live’?</td>
</tr>
<tr>
<td>Shelter</td>
<td>Is the designatum designed to protect from nature?</td>
</tr>
<tr>
<td>Comfort</td>
<td>Is the designatum felt to provide comfort?</td>
</tr>
<tr>
<td>Security</td>
<td>Is the designatum expected to provide security?</td>
</tr>
<tr>
<td>Origins</td>
<td>Is the designatum understood as a place of origin?</td>
</tr>
<tr>
<td>Belonging</td>
<td>Is the designatum held to be a place of belonging?</td>
</tr>
<tr>
<td>Possession</td>
<td>Is the designatum profiled as being owned?</td>
</tr>
<tr>
<td>Struggle</td>
<td>Is the designatum perceived as a goal in a struggle?</td>
</tr>
<tr>
<td>Building</td>
<td>Is the designatum described as something to build?</td>
</tr>
</tbody>
</table>

The analysis of such features is inherently subjective and therefore of questionable reliability. There are four responses to this important and valid criticism. First, all semantic analysis is inherently subjective. Obviously introspection and elicitation methods are subjective, but the usage-based observation of formal and objectifiable phenomena also possesses an inherently subjective dimension. Collocation studies and other corpus techniques that restrict the data to formal occurrences and co-occurrences may be objective in the actual data analysis, but the interpretation of the results of formal analysis remains entirely subjective. What it means that a given form (co)occurs more often than another form is far from obvious, especially when the exact uses of the forms in question are unknown. In contrast to such studies, the repeated and close manual nature of usage-feature analysis means that the details of the use are taken into consideration.

Second, the overt, systematic and repeated analysis of contextualised examples maximises the reliability of the analysis. The overt nature of this method means the analysis can be checked and/or repeated and multiple ‘coders’ can analyse the same data and their results compared. If this is done, a Kappa score can be used to determine the reliability of the analysis. Zeschel (2010) and Glynn (2010a) are examples of the use of such a technique.

Third, for the most part, the subjective analysis is straightforward, in that it is clear which conceptual or functional category applies to a given example.
Therefore, for the majority of examples, the analysis, although subjective, is reliable.

Fourth, the results are modelled using statistical analysis. This arguably lends a degree of objectivity to the interpretation of the results. Not only does statistical analysis tolerate a certain amount of ‘noise’ in the data, the reliability of the results can be ascertained using predictive modelling (cf. Glynn 2014a) Although this does not actually add objectivity to the analysis, it does allow a means for testing accuracy. There is a strong argument that, in semantic analysis, accuracy is more important than objectivity, which is ultimately impossible. If a subjective analysis is found to be able to predict natural language use, then this adds to the argument that the original analysis, though subjective, was accurate. Let us now consider each of the features.

The first feature is that of ‘lodging’. This category is quite straightforward and distinguishes designata that are houses or lean-tos and cabins from homelands and hometowns and streets of one’s childhood. It is exemplified in (9).

(9) Some guys they just give up living. And start dying little by little, piece by piece. Some guys come home from work and wash up. And go racin’ in the street.
   (Springsteen, Racing in the Street)

‘Shelter’ is equally straightforward and is used to identify examples where protection from the elements or nature is important. Although one would expect it to be distinctly associated with frontier literature, where the dangers of nature were real and ever-present, it is interesting, that this feature of the concept is also important in the 20th century examples. Perhaps this is because of the homelessness caused by the Great Depression about which Guthrie wrote and the hardships of urban youth culture, which Springsteen treats metaphorically as life on the streets.

(10) I was trying to make it home through the forest before the darkness falls.
   (Springsteen, My Father’s House)

The semantic feature of ‘comfort’ is slightly more subjective since it is largely determined through speculation based on context. In examples where it was not reasonably clear that this dimension of the concept of home was profiled, the feature was not annotated. Example (11) is representative of this feature.

(11) Still, there was a smiling expression of good-humor in his happy countenance, that was created by the thoughts of home and a Christmas fireside,
with its Christmas frolics.

(Cooper, *The Pioneers*)

In example (12), *home* is used in contrast to the tent in a battlefield encampment. Although *home* here surely entails comfort, we can also infer that it represents a place of ‘security’ away from the war. In both the 19th and 20th century data, there is a substantial number of examples that concern coming home or being away from home because of war but also of examples the world is depicted as a dangerous place, in contrast to *home*, which is safe.

(12) Lord, squatter, when I was a man in the pride and strength of my days, I have looked in at the tent door of the enemy, and they sleeping, ay, and dreaming too, of being at *home* and in peace!

(Cooper, *The Prairie*)

The next semantic feature is that of ‘origins’. An important, though quite marked conceptual profile is where the principle *designatum* is one’s national origin (in terms of migration) or one’s hometown / streets. This feature is surprisingly common with over 100 occurrences and is reasonably simple to identify. Examples (13a)–(13b) are representative.

(13) a. At sixteen she quit high school to make her fortune in the promised land. She got a job behind the counter in an all night hamburger stand. She wrote faithfully *home* to mama.

(Springsteen, *Big Things One Day Come*)

b. ... as this lady, a younger sister of their deceased mother, had left her paternal *home*, in the colony of Virginia ...

(Cooper, *The Spy: A Tale of the Neutral Ground*)

The most difficult feature to identify is ‘belonging’. This is designed to capture the very abstract emotional attachment between an experiencer and the *designatum* conceptualised as home. Great care was taken to restrict the annotation to instances where this feature was clearly profiled. However, the subjective nature of the category warrants caution in interpreting results based upon it, especially since it is imaginably quite a ‘central’ element to the conceptualisation of *home*. Examples (14a) and (14b) are typical.

(14) a. Wherever you may roam. You’ll never find what you left behind. Your loved ones and your *home*.

(Guthrie, *Ramblin’ Reckless Hobo Letra*)
b. ... but the truth is, their houses are floating ones, and their \textit{home} is on the ocean.

(Thoreau, \textit{Cape Cod})

Although not rare, the semantic feature of ‘possession’ was not particularly common, only 56 occurrences being identified for the target concept of \textit{home}. Typically, it is associated with the themes of land squatting and repossession / mortgage foreclosure. Its identification was straightforward as can be seen in example (15). Typically, it was the absence of possession that was profiled.

(15) a. Rich man took my \textit{home} and drove me from my door.

(Guthrie, \textit{I Ain’t Got No Home})

The feature of ‘struggle’ is also a highly subjective feature to identify yet surprisingly important with 104 occurrences. The notion of struggle included the concepts of fighting and winning and should not be understood as necessarily linked to hardship. Whether this is merely a result of the genres that make up the data set or a characteristic of the American concept of \textit{home} cannot be determined, but it is surely a result that warrants further investigation. Consider examples (16a) and (16b).

(16) a. Now I was young and pretty on the mean streets of the city. And I fought to make ‘em my \textit{home}.

(Springsteen, \textit{When Your Alone})

b. ... the cry of rapid conquest of the wilderness. We have so far won our national \textit{home}, wrested from it its first rich treasures ...

(Turner, \textit{The Frontier in American History})

The final semantic feature of the actual conceptual profiling of \textit{home} is termed ‘building’, exemplified in (17a)–(17c). This feature is straightforward and identifies instances where the building of the home, whether literal or figurative, plays a role in the conceptualisation.

(17) a. Lincoln represents rather the pioneer folk who entered the forest of the great Northwest to chop out a \textit{home}, to build up their fortunes in the midst of a continually ascending industrial movement.

(Turner, \textit{The Frontier in American History})

b. We can spend our lives in love. You’re a hesitating beauty Nora Lee. We can build a house and \textit{home}.

(Guthrie, \textit{Hesitating Beauty})
c. Wish me luck my lovely, I’ll send for you when I can. And we’ll make our home in the American land. 
(Springsteen, *American Land*)

3 Results and Interpretation

In order to understand the relative associations between the different semantic features and the conceptual structures, we employ two exploratory multivariate techniques. Firstly, we cluster the examples, using hierarchical cluster analysis. This allows us to check whether the semantic features profile the source concepts identified and not the stylistic variation between the two genres. Secondly, the data are submitted to a binary correspondence analysis. This reveals what associations are causing the clustering in the previous analysis. The systematicity of the data behaviour is explained, and any differences and similarities between the two periods identified. A third step, submits the same data to a multiple correspondence analysis in order to look for relations between the semantic features themselves. The results of this analysis are, in turn, clustered in order to determine the underlying structure of the results.

3.1 Clustering of concepts relative to semantic features

The first step is to determine how the different source concepts in question cluster, relative to the semantic features and the two periods. In other words, how do the semantic features group the concepts across the 19th and 20th century datasets? If we were to find that the concepts are clustered into two groups, 19th and 20th century, then it would be likely that the semantic features are interacting with the stylistic differences and cannot be used to describe the conceptual variation. Figure 1, below, is a dendrogram of a hierarchical agglomerative cluster analysis with multiscale bootstrap resampling.5 The distance matrix employed is the Euclidean, which is the simplest and most neutral. The agglomerating method is Ward, which is standard for small sample (Divjak and Fieller 2014).

In Figure 1, the numbers under the branches indicate the order of clustering and the numbers above are the bootstrapped confidence scores. The number to the left (au) is an unbiased probability, calculated with multiscale boot-

5 Cluster analysis performed with the R package pvclust (Suzuki and Shimodaira 2011).
Fig. 1: Cluster analysis of century and concept relative to semantics.

strap resampling and the number to the right a standard bootstrap probability. The former is argued to be more accurate (Shimodaira 2004). The bootstrapped estimated $p$-values are all high, especially considering the number of semantic features and the small sample size.

Intuitively, two clusters, 19C and 20C HOUSE and 19C and 20C NATION appear informative. That these concepts group together across the two periods demonstrates that, at least for these concepts, the semantic features are not primarily interacting with stylistic differences between the genres. In contrast to this, the cluster of 19C LAND and 19C PLACE could be argued to be a result of stylistic similarity. However, given that the 20C LAND and 20C PLACE show no evidence of clustering along genre lines and that we expect variation between the two periods, we can conclude, with some confidence, that the possibility that the semantic features merely identify stylistic differences between the two periods is not the case.

If we accept this, we have a first result. The conceptualisation of HOME as HOUSE and as a NATION has not changed over the last 200 hundred years, yet conceptualisation of LAND and ABSTRACT PLACE as HOME may have. We can now investigate these possibilities by examining how the different semantic features cluster the concepts relative to period.
[3.2 Correspondences between concepts and semantic features]

Correspondence analysis is an exploratory multivariate technique that identifies associations in complex data (Glynn 2014b). We use it here in an attempt to reveal what causes the clustering revealed in Figure 1 and, in doing so, we obtain a semantic profile of each of the conceptualisations relative to period. Before we interpret the biplot presented in Figure 2, we need to determine if the analysis is stable and if the two-dimensional representation is capable of capturing the interactions in the data. Consider the scree plot of the analysis presented, above, in table 2.

We see in the scree plot that the first two dimensions, those visualised, accurately represent 76% of the complexity (inertia). The score represents a reliable result. However, note that there is no clear ‘elbow’ in the scree plot and that the third dimension, not included in the visualisation, would contribute another 13.4% to the explanation of the behaviour of the data. This suggests that two-dimensions are not entirely sufficient to represent the behaviour of the data. For this reason, some care must be taken in the interpretation of the results.

This correspondence analysis is based on 21 mathematical dimensions, corresponding to all the semantic features concerned, minus one (the scree plot, above, includes only the first seven). The analysis calculates the contribution of each of the semantic feature dimensions to the first two axes – visualised dimensions. In other words, it quantifies how important a feature is in explaining the behaviour of, or the structuring of, the data. These values are

---

6 Scree plot produced using R package ca (Nenadić and Greenacre 2007).
also calculated for the concepts. Based on these values and the overall analysis, it is possible to calculate the accuracy, or quality, of the representation of each data point of the plot. Using the ca package in R (Nenadic and Greenacre 2007), these quality scores were calculated, each score out of 1000. Data points with scores lower than 500 should be treated with caution (Greenacre 2007). For practical reasons, the scores are not presented, but all of the concept data points obtained quality scores over 500 save 19C HOUSE and 20C HOUSE. For the semantic features, the ‘shelter’ and ‘belonging’ features both scored between 400 and 500 and the ‘struggle’ data points scored beneath 200.

Having established that the representation is reasonably stable and identified which data points could be misleading, we can interpret the results of the analysis visualised in Figure 2. Three data points, 19C LAND, 19C HOUSE and 20C PLACE, dominate the structure of the plot in terms of contribution, indicated by the size of the ‘bubble’ identifying the data point. The contributions of 20C HOUSE, 19C HOUSE and the semantic feature of ‘lodging’ are also important.

7 The correspondence analyses in Figures 2 and 3 and the various analyses in Figure 4 were performed with the R package FactoMineR (Husson et al. 2012).
The position of 20C PLACE, close to the y-axis, means that it is associated with all the features in the bottom half of the plot. However, given that the majority of the features in the left bottom quadrant contribute little to the structuring of the data (they are close to the x-axis, thus not strongly associated with the data space in the bottom of the plot), we should be careful interpreting any strong degree of correlation. Nevertheless, it is clear that the semantic features clustering broadly around 20C PLACE are characteristic of this concept. Perhaps most importantly, it must be noted that 19C PLACE groups clearly with the same set of semantic features as 20C PLACE, suggesting that the two concepts have not changed over the two centuries.

In the same quadrant, we have 19C and 20C NATION, sharing the association of the semantic features with the 19C and 20C PLACE data points. This appears to be a clear result showing that the semantic profiles of PLACE and NATION are extremely similar and that both have remained largely constant over the two periods. However, as we will see below, this particular pattern may be misleading.

The top right-hand quadrant is dominated by 19C and 20C HOUSE. This set of associations is surely stable. The position of 20C HOUSE on the x-axis means that the features in the bottom left quadrant are associated with it, in contrast to 19C HOUSE, which lies in the centre of the quadrant. Although the differences are small, they are intuitively sound: 19C HOUSE being associated with ‘shelter’ and ‘lodging’ more than the 20C, which is distinctly associated with ‘security’ and ‘struggle’. Given the urban – rural difference between the centuries and that the feature ‘shelter’ concerned protection from the wilderness contrary to ‘security’, which was understood as abstract emotional security, this kind of difference is to be expected.

The top right quadrant is clearly dominated by 19C LAND. Note, however that 20C LAND lies in the centre of the quadrant, even if its contribution is minimal. The concept of LAND, in both centuries is associated with the semantic features of ‘building’ and ‘lack of origins’. It is the second feature that makes it distinct from NATION. It appears that the concept of LAND is associated with uses where one’s ‘origins’ are lost or unattainable, in contrast to ABSTRACT PLACE, which is associated with returning to one’s ‘origins’ and with the sense of ‘belonging’ evoked by this.

These results paint a clear picture of the concept of HOME, based on the figurative uses of the single lexeme home. It seems that the four conceptualisations are reasonably stable over the 200 hundred years, though certain differences do appear. However, the cluster analysis suggested a more complex picture and one must be cautious with binary correspondence analysis for such complex data. The next section reports the results of a multiple correspondence
analysis, which reveals that, although the overall findings are accurate, there is perhaps more complexity than the binary correspondence analysis would suggest.
3.3 Semantic map of the diachronic conceptual variation

Multiple correspondence analysis follows the same principle as binary correspondence analysis, save that instead of stacking, or concatenating, all the semantic features into a single factor, they are treated as separate and independent factors. The result is more complex and less reliable, but allows us to consider how different semantic features might interact between themselves and not just in relation to the concept.

The same data, submitted to a multiple correspondence analysis produces a reasonably stable result. Normally, the explained inertia scores of the first two-dimensions are not interpretable in multiple correspondence analysis. However, Greenacre (2007: 145) has proposed an algorithm that produces interpretable scores, although with some caution. Using the adjusted algorithm, 56% of the inertia is explained. The scree plot below, in table 3, gives a dimension breakdown for the explained inertia of the analysis visualised in Figure 3.8

Greenacre (2007) has also proposed another method for estimating explained inertia, which he terms the ‘joint’ method. This method deletes the uninformative bi-rows from the calculation. Using this method, 64.7% of the variation is accounted for. Although both scores are low, relative to the binary correspondence analysis, they still represent interpretable results. The quality

<table>
<thead>
<tr>
<th>dim</th>
<th>value</th>
<th>%</th>
<th>cum%</th>
<th>scree plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.034778</td>
<td>43.0</td>
<td>43.0</td>
<td>****************************</td>
</tr>
<tr>
<td>2</td>
<td>0.010487</td>
<td>13.0</td>
<td>13.0</td>
<td>******</td>
</tr>
<tr>
<td>3</td>
<td>0.00548</td>
<td>9.3</td>
<td>65.3</td>
<td>*****</td>
</tr>
<tr>
<td>4</td>
<td>0.005985</td>
<td>7.4</td>
<td>72.7</td>
<td>****</td>
</tr>
<tr>
<td>5</td>
<td>0.000864</td>
<td>1.1</td>
<td>73.8</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td>0.000395</td>
<td>0.5</td>
<td>74.3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.0e-050</td>
<td>0.0</td>
<td>74.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3e-06000</td>
<td>0.0</td>
<td>74.3</td>
<td></td>
</tr>
</tbody>
</table>

8 Note that the correspondence analysis was performed in both the FactoMineR (Husson et al. 2012) and ca (Nenadić and Greenacre 2007) packages. The numerical summaries, quality scores and scree plot were produced using ca and Greenacre’s (2007) ‘adjusted’ method, where the biplot was produced using a standard Burt matrix the FactoMineR package. There was no noticeable difference in the plots produced by the Burt and ‘adjusted’ correspondence analyses. The FactoMineR package was used for the biplot because of its superior graphics options.
Fig. 3: Multiple correspondence analysis of concept-period and semantic features.
scores for the concepts 20C LAND and 20 NATION were still both beneath 500. Also the semantic features of ‘non-comfort’, ‘security’ and ‘struggle’ re-appear as questionable and should be interpreted with caution.

Figure 3 presents the results of the multiple correspondence analysis of concept-period and the full set of semantic features. Firstly note that the overall structure of the data is maintained. The source concept HOUSE is still clearly associated with ‘comfort’ and ‘lodging’ and the 19th century and 20th century data points are clearly sharing the same associations. Note that 20C HOUSE still lies on the x-axis and, therefore, is also associated with ‘struggle’ and ‘possession’. Given that the 19C HOUSE lies in the centre of the bottom-left quadrant, ‘struggle’ and ‘possession’ are distinctively 20th century. We obtained low quality scores for ‘struggle’ in both correspondence analyses, but it appears quite stable in relation to 20C HOUSE.

The source concept LAND also appears consistent with the previous analysis, although it should be noted that the association of the 19th century data point and the semantic feature of ‘building’ is the anchor for this clustering and that the clear association of 20C LAND appears to be drifting towards the x-axis. This suggests it is not distinctly associated with any of the features on the right side of the plot.

It is this general spread of features across the left-hand side of the plot that brings us to the major difference between the two analyses. Instead of three distinct semantic conceptual clusters, we have a continuum from 20C PLACE, strongly structuring the data at the top on the y-axis, across to 19C LAND in the bottom right-hand quadrant. It seems that these two concepts are distinct and that the other concepts are ‘floating’ between them. The contribution of the non-profiling of ‘lodging’ is, obviously, common to all these concepts and it could be that this feature is causing otherwise distinct clusters to appear associated. The question is, does patterning on the right-hand side of the plot represent two or three semantico-conceptual structures? A more straightforward way of asking this question is: does 20C PLACE and, perhaps, 20C NATION represent, quantitatively, a distinct pattern and, therefore, diachronic variation in the overall conceptual profile of HOME.

The binary correspondence analysis in Figure 2 revealed what was a clear and intuitively sound result. There was no noticeable variation between the two periods and the four concepts were structured by three semantic profiles as HOUSE, LAND, and PLACE-NATION. However, the multiple correspondence analysis reveals the possibility of a more complex picture. By looking at the interaction of the semantic features, we see that 20C PLACE, in association with ‘no comfort’ and ‘no security’, is distinct from the LAND-PLACE semantic profile and the association’s contribution to the overall structure of the data is important.
In order to determine whether the behaviour of data can be best explained as three or four structures, we can return to the cluster analysis, presented in section 2.1. Firstly, the data with which the binary correspondence analysis was performed are submitted to a $k$-mediod cluster analysis.\textsuperscript{9} Unlike the hierarchical clustering in Figure 1, $k$-mediod clusters the data with a pre-determined number of clusters. In our case, we have two possibilities – three clusters or four clusters. If we run two $k$-mediod analyses and compare the results, we can use quantitative measures to determine which clustering better explains the data.

The results of the $k$-means clustering confirm the subjective interpretation of Figure 1 and Figure 2. A silhouette coefficient measure was used to compare the clusterings, and a three-way cluster better explains the data than a four-way clustering.\textsuperscript{10} However, several important points must be made. First, these $k$-mediod clusterings are based on the stacked arrangement of the data employed in the hierarchical cluster and binary correspondence analyses. Therefore, this clustering solution tells us nothing directly about the results in Figure 3, it only confirms our interpretation of the binary analysis. Secondly, neither of the silhouette coefficient scores was high and there was not a large difference between them. The scale for the silhouette coefficient measure is: $< 0.25$ no substantial structure found; 0.26–0.50 structure is found but it is weak; 0.51–0.70 a reasonable structure identified; 0.71–1.0 a strong structure identified (UNESCO 2013). The three-way cluster produced a score of 0.52, just above the rule of thumb for a stable structure, but the four-way cluster produced a silhouette coefficient score of 0.50, also right on the cusp and only fractionally worse than the three-way cluster. Moreover, the four-way cluster identified 20C \textsc{place} as outside the general clustering of the examples with a poor individual score. In fact, it is the 20C \textsc{place} score that brings the entire silhouette coefficient score below the 0.50 threshold. In other words, although the three-way cluster is the best, the difference between the two is the behaviour of 20C \textsc{place} making it difficult to identify structures across the entire dataset. It could be that 20C \textsc{place} is so varied that it is semantically hyperonymic to the other concepts and, therefore, resists categorisation or it could be that there is something going on between the semantic features that the binary analysis is missing. It is precisely in such a situation that multiple correspondence analysis might offer an explanation.

\textsuperscript{9} $k$-mediod analysis performed with R package cluster (Maechler et al. 2012).
\textsuperscript{10} Dey et al. (2011) explain the silhouette coefficient measure. In line with the hierarchical cluster analysis, the $k$-means analysis was done using the Euclidean distance matrix.
Figure 4 presents the results of a clustering of the output of the multiple correspondence analysis. The top left plot shows the clustering of the individual examples in the correspondence analysis. The dark line through the centre is the $k$-means suggested cut. Note that the automated suggestion of the $k$-means analysis of the output of a multiple correspondence analysis is now a four-way clustering. The plot on the top right is a three-dimensional depiction of the clustering of the data on a biplot. It allows us to see how the cluster analysis is dividing up the data points in the multiple correspondence analysis. The bottom-left plot is the factor map, a colour depiction of how the cluster analysis has identified the structuring of the data into 4 clusters (here factors). The bottom-right plot is a duplication of the plot in Figure 3. It is added to aid in the interpretation of the clustering.

Fig. 4: Hierarchical cluster analysis of multiple correspondence results.
From this, it would appear that our interpretation of the multiple correspondence analysis in Figure 3 is accurate. The concept of abstract place in the 20th century represents a distinct pattern characterised by lack of ‘security’ and lack of ‘comfort’. In this, it has split off from the cluster of place-nation identified in the binary correspondence analysis, leaving a cluster of 19C place, 20C nation and 20C nation, on the one hand, and a cluster of 19C land and 20C land, on the other.

4 Summary

Four basic source concepts for home were found. These include home as a house, home as a land, home as a nation, and home as an abstract place. Three basic semantic profiles of the concept of home were identified grouping place and nation together. These were found to be reasonably stable across the 19th and 20th century datasets. However, there is strong evidence that home as an abstract place appears to be emerging as a distinct semantic profile of the concept home in the 20th century. If future research confirms this pattern, then a reasonable interpretation would be that as society becomes more mobile, both socially and in terms of physical locations, it is reasonable that the concept of home would shift from concrete sources such as house, land, and nation to a more abstract and emotionally constructed space. Despite this intuitively reasonable interpretation of the results, due to the narrow sample, restricted to five authors from two genres, any such interpretation remains speculative until a broader and more representative sample can be examined.

The aim of the paper was to demonstrate the feasibility of the multivariate usage-feature method for the description of conceptual structures. Although the sample was too small to permit confirmatory statistical analysis, the principle of usage-feature analysis / profile-based analysis has been demonstrated to adequately capture the abstract structures typical of conceptual analysis. Moreover, it was shown that usage-feature analysis enables a quantification of the phenomena in question, permitting the application of multivariate statistics. The ability of multivariate analysis to explore the complex nature of the data and identify language patterns, sensitive to social variation, was established. It is hoped that such methodological approaches will lead to the development of an analytical apparatus that identifies usage-based, rather than idealised, cognitive models.
References


