LINGUISTIC ENTRENCHMENT AND DIVERGENT CONCEPTUALIZATION IN ONLINE DISCURSIVE COMMUNITIES

by

RAGHAV SHARMA

Submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Cognitive Science

CASE WESTERN RESERVE UNIVERSITY

August, 2022

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis of

Raghav Sharma

candidate for the degree of **Master of Arts***.

Committee Chair

Dr. Vera Tobin

Committee Member

Dr. Todd Oakley

Committee Member

Dr. Mark Turner

Date of Defense

04/27/2022

*We also certify that written approval has been obtained for any proprietary material contained therein.

Table of Contents

FIGURE LIST

TABLE LIST

Linguistic Entrenchment and Divergent Conceptualization in Online Discursive Communities

Abstract

by

RAGHAV SHARMA

Given the role of distributional semantics in child language acquisition, adult linguistic development, and the conceptualization of abstract entities, the present investigation seeks to explore if the variable frequency of linguistic utterances across clusters of users in a social network can be correlated with divergent interpretations of an ostensibly shared concept within a discursive community. Are differing rates of linguistic entrenchment within a community a marker of divergent conceptualization? To begin to address this question, this present pilot study details the socio-cognitive processes underlying entrenchment of language and concepts, and develops a method for studying divergent conceptualization in the online social media network Twitter.

1. Introduction

Vast majorities of Americans consistently rank the nuclear weapons program of Iran as a top threat to U.S. and global security. A 2006 poll (Gallup, Inc, 2006) showed 80% of Americans expressing the belief that Iran was developing nuclear weapons, and by 2009 (CNN/Opinion Research Corp, 2009) this had increased to 88%. Similar majorities of Americans express concern about the potential for an Iranian nuclear weapon, as seen in a 2019 poll (Hill-HarrisX, 2019) which found 84% of Americans rated the prospect of a nuclear-armed Iran as “very” or “somewhat” concerning -- more so than Turkish military action in Syria, North Korean ballistic missile testing, or threats to domestic election integrity.

Complicating this picture, however, are repeated assertions by U.S. intelligence authorities that the Iranian government is not pursuing a nuclear weapon. National Intelligence Estimates -- joint reports from 16 U.S. federal intelligence agencies which represent the “[intelligence community]'s most authoritative written judgments on national security issues” (Rosenbach and Peritz, 2009) -- presented in 2007 and 2012 claim that Iran is not pursuing a nuclear weapon. As did former Director of National Intelligence James Clapper in a February 2015 testimony before the U.S. Congress, who said “as far as we know, [Ayatollah Khamenei]’s not made the decision to go for a nuclear weapon.” (Zenko, 2015). This phantom Iranian nuclear weapons program is one example among many of a mismatch between public perception and reality. Illusory beliefs such as this can be fruitful for investigating the social-cognitive processes which contribute to shaping public opinion more generally.

In this study, I attempt to understand more about where this mismatch comes from, how it spreads, and to what extent patterns in linguistic expression reinforce and/or disseminate semantic associations that contribute to divergent conceptualization within a discursive community. To that end, I apply perspectives from cognitive and corpus linguistics to a historical dataset of Tweets generated at key moments of public discourse on the Joint Comprehensive Plan of Action, or "Iran Deal." This project is a pilot study in two senses. The data and analysis provide initial, preliminary findings about the central research question: whether participation in one or another of the discursive communities under consideration has driven divergent conceptualizations of the "nuclear program" of Iran. In addition, the study serves as a pilot of these specific tools and methods, testing their utility for addressing empirical questions of this type. My report will address study outcomes with respect to both of these questions.

First, however, we must establish the utility of this particular discourse for the present investigation. Given the role of illusion in the study of cognition and consciousness more generally, we might expect that systematic misapprehensions among a political body will be useful for understanding the collective and agentive dynamics of meaning-making. Given the influence of frequency and sociopragmatic context on language acquisition at every level of linguistic development, from childhood learning to adult encounters with new words and phrases, we might expect online corpora with social metadata (such as Twitter, Reddit, and other Web 2.0 networks based on user-generated content) to offer a rich source for exploratory and confirmatory analysis. And so, given these expectations, we can ask how linguistic entrenchment in online discursive communities relates to divergent conceptualization of an ostensibly shared concept.

2. Language and the Emergence of False Beliefs

The modern study of individual subjective experience has often benefited from inquiry into hallucinations, illusions, and other non-veridical perceptions. Telles-Correia et al. (2015), in a historical review of hallucination and related concepts, describe various late-20th-century definitions of "illusion" in psychology and philosophy as entailing "falsifications of the perception of real objects" or "pathologically altered real perceptions." Both hallucination and illusion are part of a cluster of concepts used in philosophy of mind to illustrate and complicate everyday cognitive processes like perception and consciousness. Illusion in particular serves as a paradigmatic form of misrepresentation, deployed by philosophers for millennia (Glenney & Silva, 2019) to investigate the nature of representation itself.

Crane and French (2021) describe the "argument from illusion" as one of the principal objections to the direct realist notion of perception, the supposedly intuitive understanding that "perceptual experiences are direct perceptual presentations of ordinary [mind-independent] objects." The possibility of illusion is presented as a challenge to this folk concept of perception by philosophers ranging from Descartes and Berkeley to A.J. Ayer – the latter of whom was forcefully rebutted by J.L. Austin (see Ayer, 1940; Austin, 1962; Firth, 1964; Reynolds, 2000). Often deployed in service of neo-Cartesian arguments for sense-data as the primary objects of perception, arguments from illusion appealed to a supposed isomorphism between illusory and ordinary objects of perception and the structure of perceptual experience in the two cases. That illusions such as refraction through water or motion blur can coexist in the visual plane with ordinary mind-independent objects is taken as a challenge to direct realist accounts of perception.

In the modern era, following stage illusionist Méliès's discovery of his "stop trick" and the birth of narrative cinema, film theorists working in the Soviet montage tradition (e.g. Kuleshov, 1935; S. Eisenstein, 1949) attempted to define the specific nature of the film medium in terms of its unique cognitive affordances. Eisenstien's dialectical theory of the film image followed from illusory phenomena of the sort demonstrated by his contemporary Kuleshov. Alfred Hitchcock explained the "Kuleshov effect" in a 1964 interview:

> Now we have a close-up [of a man]. Let me show what he sees. Let's assume he saw a woman holding a baby in her arms. Now we cut back to his reaction to what he sees. And he smiles. Now, what is he as a character? He's a kindly man. He's sympathetic. Now, let's take the middle piece of film away – the woman with the child. But leave his two pieces of film as they were. Now we'll put in, uh, a piece of film of a girl in a bikini. He looks – girl in a bikini – he smiles. What is he now? He's a dirty old man. (Markle, 1964)

Just as a series of still images take on the appearance of motion through superimposition and contrast, so too is the unity of meaning achieved through juxtaposed meanings in an arrangement of images. For Kuleshov and Eisenstein and Dziga Vertov, and many post-war directors applying the cinematic theory of montage, this new filmic logic was "an opportunity to encourage and direct the whole thought process." (S. Eisenstein, 1949) This seemingly illusory phenomenon was a key insight into the semiotic structure of film.

The advent of sound film equipped psychologists and psycholinguists with more precise tools for studying multimodal interaction, as well as perceptual illusions stemming from this interaction. One such illusion is the "McGurk effect," (McGurk & MacDonald, 1976) which Tiippana (2014) describes in its best-known version as the audiovisual illusion induced "when dubbing a voice saying [b] onto a face articulating [g] results in hearing [d]." The McGurk effect is taken to indicate a unity of perception across the various sensory modalities – or, more conservatively, overlap or integration of the various sensory processes – in yet another example of illusions shedding light on cognition more generally.

Extending this investigation beyond sensation and perception into the realm of social cognition and language, Tobin (2018) examines presuppositions embedded in linguistic constructions by analyzing "illusions of knowledge" (136): instances of "erroneous" interpretation due to indeterminacy of linguistic reference that, error aside, rely on the same cognitive biases underlying "normal" linguistic reference and sentence interpretation. Like the use of illusion to profile aspects of sensation relevant to a full understanding of consciousness, the study of narrative surprise allows researchers to examine the cognitive processes underlying presupposition, which is highly salient in certain circumstances but most often invisible, even while profoundly structuring how language contributes to meaning-making. Indeed, Tobin identifies other circumstances in which illusions of knowledge are generated through presupposition, including push polls employed by political campaigns and leading questions designed to elicit specific answers from witnesses, demonstrating the broad relevance of presupposition in the relationship between language and knowledge. Most notably for the current investigation, giving a name to "entities that fail, in one way or another, properly to exist" (130) presupposes the existence of the named entity.

The cognitive dynamics of presupposition are described by Tobin using the framework of mental spaces theory (Fauconnier, 1997), in which mental spaces are "collections of small, local, structured, interconnected mental representations" intended to describe human understanding of propositional content in a more parsimonious manner than the "possible worlds" approach from 20th-century analytic philosophy (e.g. Lewis, 1969). This same framework is applied by Coulson (2006) in analyzing cultural models, mental space networks representing shared concepts of sociocultural phenomena. Framing the cognitive phenomena being investigated in terms of mental spaces allows both the account of presupposition and the account of cultural models to profile how individual speakers can exert epistemic or rhetorical influence within a discourse. Indeed, both Tobin and Coulson describe their cognitive models of meaning-making as highly susceptible to individual rhetorical intention -- whether literary, commercial, or political.

One purpose of this paper is to focus on divergent conceptualization as a useful perspective on the normal social-cognitive dynamics of understanding abstract phenomena in the world which most individuals have no direct access to, such as entities involved in foreign policy discourse. We can pose this inquiry in terms of a number of questions: are there linguistic markers of divergent conceptualization within a discursive community? If so, can the frequency and routinization of these markers serve as an indicator for conceptual development over time, structuring the interpretation of a given concept in progressively more predictable ways? As we will see, the process of entrenchment – "the degree to which the formation and activation of a cognitive unit is routinized and automated" (Schmid 2010) – bears heavily on linguistic development, language acquisition, and the conceptual basis of linguistic abstraction. Does the entrenchment of particular linguistic units suggest the entrenchment, as well, of an underlying conceptual structure – and can entrenchment of these conceptual structures reliably produce particular presuppositions?

2.1 Entrenchment in Language

The relationship between frequency of usage and lexical or semantic entrenchment is a frequent object of study in cognitive linguistics over the last quarter century (e.g. Haiman 1994; Bybee and Thompson 1997; Lupyan and Winter 2018; Tobin 2021). Earlier work by Geeraerts, Grondelaers, and Bakema (1994) operationalized a quantitative measure of entrenchment based on the frequency of particular word choices within a corpus. Entrenchment is linked to the authors' notion of "onomasiological salience," or the preponderance of a given lexical item within a lexical field (the set of linguistic units related to a given concept).

Drawing on Eleanor Rosch's prototype theory (Rosch, 1978; Green and Evans, 2006, p. 256), Geeraerts, Grondelaers, and Bakema caution against an intuitive notion which correlates onomasiological salience with the prototype structure of a category, thereby suggesting that the most salient lexical units in a given field are found at the basic level of categorization. The authors instead argue for generalizing this notion due primarily to the observation of "differences of onomasiological preference" on an intra-category level. They articulate this objection to the "basic level model" of onomasiological preference with an example:

> The basic level model contains a hypothesis about alternative categorizations of referents: if a particular referent (a particular piece of clothing) can be alternatively categorized as a garment, a skirt, or a wrap-around skirt, the choice will be preferentially made for the basic level category "skirt". But analogously, if a particular referent can be alternatively categorized as a wrap-around skirt or a miniskirt, there could just as well be a preferential choice: when you encounter something that is both a wrap-around skirt and a miniskirt, what is the most natural way of naming that referent? (Geeraerts, Grondelaers, and Bakema, 1994, p. 137)

This "generalized onomasiological salience" is then equated with entrenchment.

Geeraerts, Grondelaers, and Bakema are primarily concerned with quantifying lexical entrenchment, or onomasiological salience within a given lexical field, as a means of better understanding lexical preference. Their conclusion from their findings is that entrenchment does indeed influence lexical preference. Thus, the preponderance of individual lexical units is influenced (among other things) by their degree of entrenchment within the set of words used to discuss a given concept.

Building on earlier foundational work (Hooper, 1976; Bybee 1985; Haiman, 1994) on the role of entrenchment in language, Bybee and Thompson (1997) extend the analysis of the role of lexical frequency in entrenchment. The impact of individual utterances, accumulated over time, is made explicit in the authors' claim that "each token of use [has] a potential effect" (379) on the overall representation of a meaning in a given linguistic form. Among other impacts of lexical frequency on syntax, the authors highlight both a "reduction" and "conserving" effect from increased frequency of lexical tokens.

The "Reduction Effect" refers to phonetic, syntactic, and semantic shortening of high frequency tokens. One example on a syntactic level is the grammaticization of commonly used passive constructions paired with infinitives, resulting in the shortening of (e.g.) “going to” and “have to” into “gonna” and “hafta.” Bybee and Thompson also identify a “Conserving Effect” of token frequency, which inhibits morphological change of linguistic tokens which occur most frequently. According to the authors, the Conserving Effect of frequency in syntax involves high frequency tokens “[taking] on a life of their own, and resist[ing] change on the basis of newer productive patterns.” (381) As an example, they cite the morphological phenomenon of lower frequency verbs “regularizing” to more conventional morphology compared to high frequency token verbs sharing a single alternation pattern – e.g. “weep” and “creep” regularize into “weeped” and “creeped” while “keep” and “sleep” maintain their morphological irregularity when changing tense into “kept” and “slept.”

Bybee and Thompson appeal to earlier research (Haiman, 1994; Bybee 1985, 1997) which suggests that these representational tendencies are products of more general features of human and animal cognition. For the Reduction Effect, this is due to the increased efficiency of “any motor activity that is repeated often” (Bybee and Thompson, 1997). The Conserving Effect arises due to the strengthening in memory of the representation of a given lexical form as it is repeated more frequently – “high frequency forms" exhibit “faster lexical access.” On a semantic level, the authors connect these entrenchment-related phenomena to Haiman’s (1994) theory of “ritualization” – “varieties of change which are brought about through routine repetition” – as central to the development of language over time. Haiman analogizes ritualization in language to similar entrenchments of human and animal behavior through repetition. This appeal to phylogeny demonstrates that the psychological processes enabling entrenchment precede any specifically linguistic faculty.

2.2 Entrenchment in Linguistic Development

Basing historical theories of language change in general cognitive functions lines up with the phylogenetic and ontogenetic account of linguistic development presented in Tomasello's *Constructing a Language* (2003), according to which language learning is made possible by two suites of cognitive functions with a clear evolutionary lineage in human and primate behavior: a set of social-cognitive functions on one hand and a faculty for pattern-finding and abstraction on the other. This is distinct from the standpoint of universal grammarians (e.g. Chomsky 1965) who suggest that knowledge of any given language is an instantiation of a genetic faculty for language characteristic of humans. From the standpoint of construction grammarians such as Tomasello (himself building on e.g. Fillmore, Kaye, and O'Connor, 1988; Goldberg, 1995; Croft, 2001), the normative endpoint of linguistic development entails precisely the sort of learning described by Bybee and Haiman as resulting from processes not exclusive to humans at all.

For construction grammarians, knowledge of a language entails applying these learning processes to a robust network of form-meaning pairs (called *constructions*). Goldberg (2003) highlights a number of key tenets distinguishing the construction grammar theory from its 20th-century counterpart, including among others: (1) language at every level (morphemes, words, recurring grammatical patterns) can be described in terms of form-meaning pairs called constructions which merge the perceptible aspect of the word with a discourse or semantic function; (2) constructions are learned exclusively through environmental input applied to general (non-language-specific) cognitive mechanisms; (3) human knowledge of language is just knowledge of constructions, with varying degrees of abstractness.

Outlining an ontogenetically and evolutionary coherent theory of linguistic development, Tomasello (2003) shows how socio-cognitive faculties and pattern-finding abilities characteristic of young children can explain away the universal grammarian “poverty of the stimulus” hypothesis, which proposes the existence of an innate language faculty to explain how children can develop robust mastery over a given language despite not receiving direct exposure to its full range of well-formed sentences. For construction grammarians, however, “there is no poverty of the stimulus when a structured inventory of constructions is the adult endpoint.” (7) According to Tomasello, four basic sets of psychological processes contribute to the development of this structured inventory:

1. *intention-reading and cultural learning* to identify scenes of communicative intention in daily experience as well as the linguistic symbols employed therein;
2. *schematization and analogy* to enable abstraction from individual utterances in a child’s environment to underlying linguistic forms of which the individual utterances are instantiations;
3. *entrenchment and competition* as a means by which – in line with the Conserving Effect described by Bybee and Thompson (1997) – children learn to limit their abstractions to forms which appear with frequency in their environment;

4. *functionally-based distributional analysis* to enable the sorting and grouping of linguistic forms according to their communicative function.

These four processes alone, on the constructionist account, can explain linguistic development "from here to there" – from infant babble and proto-syllables to mature adult speech – and does so at every stage of development without appeal to a metaphysically and evolutionarily dubious innate universal grammar.

As an example, Chapter 3 of *Constructing a Language* discusses the initial acquisition of words by young children. Tomasello challenges the folk concept of a "pointing-and-naming game" (43) as the primary means by which children start to associate spoken words with aspects of their environment, opting instead to explain this development in terms of the processes listed above. First, the pointing-and-naming game can hardly suffice for more abstract referents than concrete objects, the words for which nonetheless occur regularly in early childhood acquisition – words such as "go" or "of." Furthermore, even in the case of concrete objects, such as a toy car, it is unclear which aspect of the object is being profiled without knowledge of the word being used. "Toy," "car," "Volkswagen," "wheel," "drive," "vroom," could all equally stand in for the object on display in such a point-and-name game.

Tomasello develops a usage-based account of standard linguistic development from undifferentiated auditory input to proficiency with complex and abstract constructions. First at 14 months are *holophrases*, or short discrete lexical units derived from a word or phrase (e.g. *lemme-see*); then at 18 months come *pivot schemas* and *word combinations*, in which multiple aspects of an experiential scene are profiled with differing degrees of salience (e.g. *More X* and *Open box*); next at 18-20 months are *item-based constructions*, in which children identify participant roles within a scene standing in meaningful relation to one another, often centered around a verb (e.g. *X hit Y*); finally culminating in *abstract constructions* which develop throughout life, highly schematized forms which convey consistent meaning and are lexically instantiated according to a speaker's context and vocabulary (e.g. the ditransitive, sentences of the form *X VERBed Y NP*).

This capacity for linguistic abstraction develops alongside advances in children's general cognition and social cognition, particularly skills of pattern-finding and analogy as well as joint attention and shared intentionality. Most significantly, however, for suggesting the impact of sociopragmatic factors in language acquisition, the vast majority of early linguistic input does not come in the form of explicit instruction but rather "in the ongoing flow of social interaction and discourse." Tomasello emphasizes distributional analysis – "[grouping] together into paradigmatic categories linguistic items that behave in the same way" – and particularly a notion of "functionally-based distributional analysis," in which the similarity in behavior is one of pragmatic-communicative function. Entrenchment becomes one of the core mechanisms by which these distributional tendencies reproduce themselves. This is as evident at the level of phonology and individual words as it is for more abstract constructions such as recurring syntactic forms and discursive behaviors and, most relevant to the current study, concepts.

2.3 Entrenchment in Conceptualization

Indeed, the frequency and habituation characteristic of entrenchment in language remain incredibly salient to conceptual entrenchment. Entrenchment involves a routinization of mental representations for a given input, linguistic or concept, enabling its storage and increasingly automatized retrieval in long-term memory. Schmid (2010) describes a correlation between the degree of entrenchment for a concept and its frequency of activation, such that greater entrenchment implies greater frequency of activation and vice versa. This entrenchment and activation occurs in the context of a “speech community” (Schmid 2015) in which linguistic knowledge is shared, despite no two members of the community possessing identical linguistic knowledge. Individual utterances by individual community members are both informed by and partially constitutive of the developmental tendencies within a discourse.

The process of entrenchment bears heavily on how patterns of disambiguation develop within a discourse. It is common for speakers to resolve ambiguity in language through discourse rather than explicit reference. When words or other linguistic units can potentially instantiate a number of concepts, entrenchment of conceptual configurations within a discourse impacts which concepts are elicited. The delegation of information-transmission to context and inference (Piantadosi et al. 2012) is to be expected given both the imperative for efficiency in speech (Wasow 2015) as well as the pervasiveness of ambiguity at the lexical and syntactic levels in spoken language (Wasow et al. 2003). Wasow and co-authors note that many of the most used nouns, verbs, and adjectives in English can be classified under multiple grammatical categories, and many even have multiple possible interpretations within a given category. Learning how to interpret a given linguistic construction often involves learning to disambiguate through context.

As Lupyan and Winter (2018) write: "the distributional structure of language provides an enormously rich source of knowledge." In trying to understand the predominance of highly abstract referents among some of the most ubiquitous words across lexical categories, Lupyan and Winter point to a number of means by which language can provide the information needed to develop knowledge of abstract concepts. Beyond the obvious utility of conveying language for propositional knowledge, the authors point to the role of verbal labeling in category formation, for concrete meanings but especially for abstract meanings, as well as processes of statistical learning performed on the distributional structure of linguistic input.

The ability (albeit constrained) of language models to accurately represent geographic and semantic knowledge based purely on linguistic input suggests that the information underlying this knowledge is present in the structure of language. This leaves open the empirical question of whether human beings engage in such learning. Lupyan and Winter go even further and suggest that categories denoted by abstract words often "[do] not exist apart from language." Taking as an example the word "fun" and its associated abstract referent, Lupyan and Winter state:

> A person never exposed to the various ways that English speakers use this word would certainly lack the relevant word meaning. Would they nevertheless have the concept? We think not. Recall that on a traditional perspective, words are thought to map onto pre-existing concepts... But what is the pre-existing conceptual representation that fun would map onto? On our view, it is observing the same word used across many disparate contexts that helps create a category which otherwise

does not exist. We can get a hint of the kind of information linguistic experiences with the word fun conveys by examining its semantic neighbourhood in a model of distributional semantics.

Given this relationship between linguistic context and concept formation, observing distributional tendencies in how abstract concepts are developed through language might provide insight into the cognitive dynamics of conceptual entrenchment.

3. The Current Study

3.1. Cognitive Linguistics

Given this role of frequency and routinization in the entrenchment of both language and concepts, to what extent is the recurrence and co-occurrence of particular linguistic units indicative of an underlying conceptual structure? If we can presume such a relationship, can distributional analysis of constructions such as words, phrases, and semantic frames (Fillmore, 1982) guide us in our understanding of a given concept? Finally, do associations between lexical units and these potentially underlying concepts drive divergent interpretations of reported phenomena? The object of the present investigation is developing exploratory methods and potential answers for these questions.

One primary task is the identification of precise forms of conceptual structure which might be sensitive to this presupposition-generating entrenchment, and attempting to explain this sensitivity in terms of domain-general cognitive processes involved in this form of conceptual structure. Tobin's definition of irony as a viewpoint phenomenon (Tobin and Israel, 2012; Tobin 2020) and "irony attrition" as a process of routinization and entrenchment resulting in the collapse of complex viewpoint arrangements into simpler configurations leading to the loss of ironic distance from an initially ironic act (Tobin, 2021) offers a model of this approach. Due to general limits on source memory not unique to language, intermediate viewpoints in these complex viewpoint structures are more likely to be the targets of semantic-pragmatic bleaching, resulting in the collapse of ironic distance.

We can generalize this perspective on complex viewpoint arrangements to our investigation into entrenchment of other conceptual structures, and return now to the mental space networks entailed in cultural models (Coulson, 2006) and presupposition (Tobin, 2018). Cultural models are shared representations of sociocultural phenomena, and are evoked and adapted for rhetorical ends by discourse participants. The adaptation of a public cultural model for the idiosyncratic ends of an individual speaker requires dynamic and imaginative meaning construction accomplished through conceptual blending (Fauconnier and Turner, 2001). Conceptual blending theory describes how novel meaning can emerge from the integration of existing concepts, even beyond the original meaning contained in those concepts.

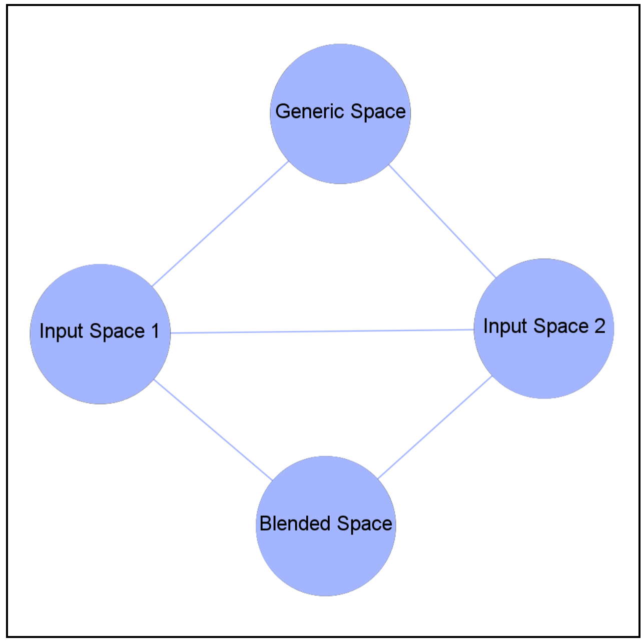

FIG. 1. A generic conceptual blending diagram, with two input spaces whose shared elements provide structure to the generic space, and whose elements are selectively profiled in the blended space arranged according to structure imported from the generic space.

Concepts being integrated within a conceptual blending network are described as input spaces -- distinct sets of mental experience organizing aspects of the world according to their relations. A conceptual blending network contains multiple input spaces as well as a generic space, representing shared elements of the structure of the inputs. The result of this network is a blended space, containing elements from both frames as well as novel elements unique to the blend.

Inputs into a conceptual blending network can include semantic frames (Fillmore, 1982), which are conceptual structures for representing knowledge of words organized according to human activity and experience. Discourse entities are represented as abstract elements whose relations are structured by the frame. These entities are instantiated as specific lexical units which evoke a given frame -- for example, the verbs “buy” and “sell” evoke the canonical “Commercial Transaction” frame. On Coulson’s account, cultural models are “a special kind of frame that deals with socially relevant topics.” When a given frame is evoked, only parts of its structure are profiled. When two or more frames are blended (Turner, 2006), the structure of the resulting blend is partially determined by the shared structure of the inputs but can also have novel properties.

Let us recall now our understanding of lexical entrenchment, and the “carving” of conceptual structure through distributional semantics within a linguistic community (Lupyan and Winter, 2018). We can expect that one factor influencing conceptual blending of cultural models is the local entrenchment of linguistic and conceptual structures. As Coulson and Oakley (2005) argue in the case of blends resulting in metaphors, whether speakers/listeners interpret a given metaphorical utterance as novel or conventional depends on its level of entrenchment in that language user’s experience. Despite sharing an identical conceptual structure, these two forms of interpretation are rooted in distinct cognitive processes: active analogical reasoning for novel metaphors and more automatic processes of retrieval for more conventional metaphors. As a given metaphorical utterance is entrenched and conventionalized, what is routinized is not just a particular interpretation of that utterance but also inferential behavior in the blended space of that metaphor – drawing not just on other linguistic input but also background knowledge, context, and more agentive inference as well – such that particular interpretations are reliably reproduced.

Moving from the mental space networks underlying conceptual blending in cultural models and metaphors to the mental space networks underlying presupposition in language, discourse, and narrative (Fauconnier, 1997; Tobin, 2018), we can see how this sort of routinized retrieval of mappings across a network results in illusions of knowledge. Taking as an example the sentence "I have to go buy some groceries for my brother," (144) Tobin demonstrates how ordinary reference often serves to introduce presuppositions into a discourse by asserting the existence of the referent, and does so in a way that is particularly resilient to refutation. By introducing new information into the common ground in a more covert manner than explicit assertion, presuppositions are able to be easily projected across an entire mental space network unless they are explicitly refuted. Thus the entrenchment of mental space configurations and their associated inferential behaviors also entails the entrenchment of any presuppositions embedded as inputs to the overall conceptual structure.

This makes presupposition a powerful tool for rhetorically-motivated speakers to drive divergent conceptualization of a cultural model in line with their rhetorical goals. But thinking back to the claim from Bybee and Thompson (1997) that each token of use has a potential impact on the process of lexical entrenchment, as well as Tobin's discussion of "presupposition accommodation" in Lewis (Lewis, 1979; Tobin, 2018), we might also suggest that the conceptual entrenchment underlying divergent conceptualization through differences in presupposition operates identically regardless of any underlying rhetorical intent by a presupposition-generating speaker. Just as in the case of metaphor conceptualization (Coulson and Oakley, 2005), the structure of the mental space network remains the same whether a presupposition is deliberately or incidentally introduced into a discourse.

3.2. Corpus Linguistics

In turning to distributional semantics as a source of information about inferential behavior in a mental space network, we require corpus linguistic methods (e.g. Stubbs, 2007; Gries, 2012, 2017) which enable the use of readily-accessible observational data and can be combined with a frame semantic analysis (Atkins, et al., 2003) to study the development of cultural models over time in a naturalistic setting. For cultural models with a political valence, there are a number of potential corpora to be analyzed for change over time. For the purposes of this investigation, the corpus of choice is Twitter, which has increased in relevance as a source of natural language data in recent years (Zottola, 2020).

Just as usage-based linguists take a maximally-empiricist approach and reject the generative claim of linguistic knowledge not acquired through experience, so too does John Sinclair (2004) – one of the founders of modern corpus linguistics – exhort researchers to "Trust the Text." In other words, both traditions emphasize the study of naturally-occurring language and both traditions let the data suggest effective categories for analysis, rather than exploring data with specific categories in mind (Stubbs, 2007). Theoretical synergy between these disciplines has lent itself to a prodigious body of corpus-assisted construction grammar research (e.g. Römer, O'Donnell, & Ellis, 2015). The present study aims to expand this fusion into investigations of novel data. If social variables in the observed network of Twitter users correlate with divergent rates and targets of linguistic entrenchment, it might suggest that a simultaneous process of divergent conceptualization is also occurring in this discursive community.

3.3. Hypotheses

The primary hypothesis being investigated in this study is the claim that divergent conceptualization in the Iran Deal discourse will correlate with the underlying social dynamics of a user's linguistic input on the Twitter app itself. In other words, this primary hypothesis (H1) suggests that the users followed by any given user will be a likely indicator of which topics a user discusses, as well as the language used to discuss that topic.

> **H1 (the clustering hypothesis): Participation in a given discursive community drives divergent conceptualization of abstract entities.**

This hypothesis could be bolstered by the validation of a number of predictions, some of which will be tested here, and some which might be the basis of future research, as discussed in Section 5.

> **P1.0: There are identifiable modular clusters within the overall network of Iran discourse on Twitter.**
>
> **P1.1: Frequency of weapons-related language will correlate with the cluster a node belongs to.**
>
> **P1.2: Frequency of energy-related language will correlate with the cluster a node belongs to.**

P1.3: For nodes belonging to the same cluster, lexical units and frames will propagate from nodes with high degree centrality to low degree centrality over the course of discussion around a given event.

An alternative hypothesis might hold that a large enough majority of public opinion coalesced around a single conceptualization (in this case, of the Iranian nuclear program as a nuclear weapons program) precludes the entrenchment of linguistic structures and inferential behaviors which might produce divergent conceptualization. This hypothesis (H2) is restated below and can also be tested with a prediction corresponding with P1.3.

H2 (the ship has sailed hypothesis): Certain conceptualizations are sufficiently cemented within a discursive network that their use does not differ significantly among groups.

P2.1: For nodes belonging to different clusters, lexical units and frames will propagate from nodes with high degree centrality to low degree centrality over the course of discussion around a given event.

4. Methods

4.1 The Objects and Structure of Twitter Data

In addition to a particular question about the relationship between linguistic and conceptual entrenchment in discourse on the Iranian nuclear program, the present investigation is an exploration into the unique affordances of the social media network Twitter for corpus and cognitive linguistics. As a constantly-expanding source of naturally-occurring (non-elicited) language data, often multi-modal and rich with metadata allowing researchers to make inferences about socio-pragmatic factors (such as geolocation, language, and interactions between speakers), Twitter is also constantly expanding in its utility for linguists (e.g. J. Eisenstein et al., 2014; Šćepanović et al., 2017; Okonski and Ferreira, 2019) as well as other social scientists using computational linguistic methods (e.g. Gaumont et al., 2018; Reyes-Menendez et al., 2020). The downside of this constant innovation in material affordances for researchers and users alike is that processes of data collection and analysis must be regularly refined and even totally redesigned resulting from changes to the data Twitter makes available.

As an example of how changes in the basic object models of Twitter (and changes to the structure of relationships between them) demand methodological sensitivity to the current state of the platform, the Twitter Developer Platform offers the following anecdote in the "metadata timeline" for their "Full-archive Search":

> For example, @Replies emerged as a user convention in 2006, but did not become a first-class object or event with 'supporting' JSON until early 2007. Accordingly, matching on @Replies in 2006 requires an examination of the Tweet body, rather than relying on the to: and in_reply_to_status_id: PowerTrack Operators.

The platform has tended towards increasing reifications of user behavior within the object model of the Tweet object itself, and the "fields" available for the v2 Twitter API's Tweet object – attributes of the Tweet object that can be optionally attached to any request – allow researchers to construct conversations over time, track patterns of retweeting and quote-tweeting (in addition to the aforementioned replies), and filter tweets by content, language, location, and more. The default fields returned with every request for a Tweet object are “id” (a string of text representing a unique numerical ID for the Tweet) and “text” (a string of text representing the body of the Tweet). Relevant fields for this study also include “created_at” (an ISO 8601 format date indicating the date and time a Tweet was generated), “conversation_id” (a string of text representing a numerical indicator linking a reply to the original Tweet it is responding to), and “author_id” (a string of text representing a unique numerical ID for the user who authored the Tweet).

Representation of Twitter users is also a major feature of the Twitter API. The user object model has three default fields returned with each request: “id” (a string of text representing a unique numerical ID for a given user), “name” (a string of text representing the display name of this user), and “username” (this user’s “handle” or screen name, often reported with the preceding “@” used to tag a user based on their username). The isomorphism between the “author_id” of a Tweet object and the “id” of a user object offers a basis for connecting the linguistic data of a Tweet’s text to the sociopragmatic inferences made possible by the affordances of the user object.

4.2 Data Collection and Analysis in Python

A number of open-source packages available in Python (version 3.8.8) enabled the collection of Tweets, replies, and users, as well as the subsequent representation and analysis of relationships between users in a network graph. First, Tweepy (version 4.8.0) was used as a container for the Twitter API, with pre-built functions making advanced use of the Requests package that traditionally retrieves responses from the Twitter API.

This allowed for the translation of data from Twitter responses into structures more suitable for processing with other, non-Twitter-specific Python packages.

The first and most broadly applicable of these more general packages is the open-source library Pandas (version 1.4.2). Pandas is built for data processing and manipulation, structuring data in tables known as dataframes which have a wide variety of uses across the data collection and processing stages. First, dataframes can be used to represent the data themselves – as in Fig. 2 below, a sample of Tweets collected from Twitter accounts affiliated with news outlets. Dataframes can also serve as an intermediary structure for writing data to or reading data from .csv, .xslx, and .json file formats, which are all popular methods of storing delimited data in tables. Finally, in conjunction with this second function, representing data in a dataframe is often a necessary step in moving from one stage of the processing pipeline to another through various packages for data representation and analysis.

	Tweet ID	Text	Author	Timestamp	Metrics
0	405833757946834944	What do you think about the Iran nuclear deal?...	759251	2013-11-27 23:01:22+00:00	{'retweet_count': 51, 'reply_count': 32, 'like...
1	404629298419466240	20 questions about the Iran nuclear deal: What...	759251	2013-11-24 15:15:17+00:00	{'retweet_count': 112, 'reply_count': 17, 'lik...
2	404581375275069440	Israeli PM Benjamin Netanyahu calls Iran nucle...	759251	2013-11-24 12:04:51+00:00	{'retweet_count': 172, 'reply_count': 61, 'lik...
3	404461797639340032	#Breaking Historic deal struck between Iran an...	759251	2013-11-24 04:09:41+00:00	{'retweet_count': 348, 'reply_count': 47, 'lik...
4	427048474333691905	Listening Post: Iran's Message at Davos Has Ee...	807095	2014-01-25 12:01:05+00:00	{'retweet_count': 37, 'reply_count': 8, 'like_...
...	...	...	...	...	...
1680	1171775432213127168	'With the firing of the biggest supporter of w...	4898091	2019-09-11 13:20:08+00:00	{'retweet_count': 7, 'reply_count': 1, 'like_c...
1681	1171698669395021824	Three Australians, two of whom also hold UK pa...	4898091	2019-09-11 08:15:06+00:00	{'retweet_count': 5, 'reply_count': 0, 'like_c...
1682	1172872883229679617	Iran's "Blue Girl" Dies After Self-Immolation ...	16935292	2019-09-14 14:01:00+00:00	{'retweet_count': 36, 'reply_count': 4, 'like_...
1683	1172687662211354624	Iran's "Blue Girl" Dies After Self-Immolation ...	16935292	2019-09-14 01:45:00+00:00	{'retweet_count': 65, 'reply_count': 13, 'like...
1684	1172155441582870529	Will U.S. & Iran Resume Talks After John "...	16935292	2019-09-12 14:30:09+00:00	{'retweet_count': 14, 'reply_count': 3, 'like_...

1685 rows × 5 columns

Fig. 2. Representation of Tweet object and various fields in Pandas dataframe.

Tweepy and Pandas can be used in conjunction to generate a dataframe of edges (connections between nodes) referred to as a "pandas_edgelist" in the syntax of the next relevant Python package, NetworkX (Hagberg et al., 2008). The edgelist (e.g. Fig. 3) is generated by inputting a list of user IDs and outputting a two-column dataframe of "sources" (the initial user IDs) and "targets" (the user ID of every user followed by the initial users). This dataframe can then be used to generate a network graph in NetworkX. NetworkX is also capable of performing various algorithmic analyses of the network, as well as some basic visualization.

	source	target
0	825143318891917313	27699644
1	825143318891917313	18823758
2	825143318891917313	920023898154065921
3	825143318891917313	415850821
4	825143318891917313	403034630

FIG 3. First five rows of edgelist representing relationship of source following target.

This edgelist, and corresponding sublists representing the "source" and "target" columns, can be saved as .csv files and serve as inputs into the Gephi network visualization software, described further in Section 4.4.

4.3 Twitter Full-Archive Querying and Retrieval

At present, the Academic Research tier of access to the Twitter API grants researchers the ability (dubbed a "Full-Archive Search") to retrieve Tweets from the entire historical archive of Twitter dating back to the first Tweet in March 2006. Tweets and users in the archive are organized according to their respective object models as outlined in Section 4.1. The Full-Archive Search is one "endpoint" among many, each representing a different source of Twitter data – including archived Tweets, users and their interrelations, and a random stream of incoming Tweets. Making requests to these endpoints returns the relevant objects – Tweets or users, for this study – in a .json file format. Other endpoints can also be used to agentically engage with Twitter by posting or deleting Tweets or engaging with other users by following, unfollowing, or blocking them.

The Tweepy package mentioned in Section 4.2 is, again, a Python wrapper around the Twitter API. Tweepy works by creating a Python iteration of the Twitter API client and contains a number of methods (re-executable chunks of code affiliated with an object, which in this case is the client) to format queries to the API endpoints and process the returned results. Each method has a number of parameters used to further refine the list of objects returned, as well as to expand the fields returned along with each user or Tweet object in the request.

In order to query the historical archive of Tweets (the aforementioned "Full-Archive Search"), Tweepy provides the .search_all_tweets() method. As its one required parameter, .search_all_tweets() takes a string variable representing the query – formatted by the user according to the operator guidelines discussed above, but inputted into the Tweepy client method and converted into the proper HTTP format for the API endpoint. Optionally, however, this Tweepy method can take a number of other parameters, which help further tailor the search to the user's needs. Relevant to the present study are parameters including "end_time" and "start_time" (which limit the search to Tweets within timeframes represented as strings containing ISO 8601 timestamps in the "YYYY-MM-DD HH:mm:ss" format measured at the UTC timezone), "max_results" (increased to the Tweepy system limit of 500 per query), and "tweet_fields" (which, again, expands the Tweet objects returned by this query with information including "author_id," "created_at," "id," "public_metrics," "referenced_tweets," and "text"). Tweets are returned as a .json from the API but are converted to a Tweepy "Response" item, which can be transformed into a .json using other open source Python libraries, but can also be transformed into Pandas Dataframes, enabling more precise processing and transferring of the data.

The next relevant Tweepy method is .get_users_following(), which allows us to generate a list of "id" strings for each user whose "id" or "username" field we input as the method's one mandatory parameter. Again, an optional "max_results" parameter facilitates bulk collection of results, in this case with a maximum of 1000 users returned per API call. By using the pagination function of the Twitter API, we can loop through a user's following list to generate a one-step "ego network" centered around a single user. This network can be represented in a Pandas DataFrame of the sort seen in Fig. 3, as a list of source and target nodes, and the NetworkX package can generate a graph of this network as edges between nodes representing a one-way follower relationship from source to target.

The graph can then have various network structure analyses applied to it by graph object methods built into NetworkX. This includes, for our purposes, the Louvain method of community detection (Blondel et al., 2008). Nodes can also be analyzed for their degree, or total number of connections to other nodes, and ranked according to degree. Low degree nodes can be removed from the network to facilitate analysis and visualization. Lists of the nodes (classified with their Louvain group and ranked by degree as seen in Fig. 4) and edges from this graph can then serve as the inputs for more comprehensive analysis and visualization, as demonstrated in Sections 4.4 and 4.6.

0	Id	degree	group
0	559906863	1421	1
1	225117039	1318	3
2	778045033828057092	979	4
3	23757702	843	1
4	1084582921682518017	836	3

FIG 4. The head (or first 5 rows) of a Pandas DataFrame representing the 5 highest-degree nodes in a subset of our network, along with their degree and cluster in the Louvain analysis.

4.4 Gephi

By converting the NetworkX graph object into two Pandas Dataframes representing its nodes and edges, and then converting these DataFrames into .csv files using the .to_csv() method built into the DataFrame object, we have distilled the structure of the graph into a ubiquitous file format that can serve as the input to any number of graph analysis and visualization processes. While NetworkX is capable of robust network analysis for graphs under a certain size, its visualization capabilities – working in conjunction with the ubiquitous Python data visualization library Matplotlib – are rather sparse. The open-source graph and network analysis software Gephi (Bastian, Heymann, and Jacomy, 2009) is a useful tool both for visualizing networks and for exploring their underlying structure.

The interface for a Gephi 0.9.2 project is split across three tabs. Beginning with the “Data Laboratory,” users can import .csv files of the nodes and edges of the graph to be visualized. The columns of the node list can be inspected upon import to ensure the fidelity of the transfer, and various transformations and filtering operations can be performed in this tab. After importing and organizing our nodes and edges, we can manipulate the visualization of this graph in the “Overview” tab. The “Overview” tab also offers a number of filters to dynamically re-visualize the graph, as well as summary statistics. Finally, after the structure of the graph is finalized using the low-resolution (but rapidly generated) images of the “Overview” tab as a model, the graph can be generated in higher resolution and exported as a .png or .pdf file in the “Preview” tab. A number of visual adjustments such as label fonts and the design of nodes and edges can also be made here.

4.5 STM

Having now analyzed the structure of our network, we must bridge the gap from these statistical and topological observations to linguistic hypothesis-testing using the related text data. An open-source library called “STM” (Roberts et al., 2019) for the statistical computing and visualization programming language R (R Core Team, 2021) serves as precisely this sort of bridge. The STM, or *structural topic model*, is a distinct approach to the task of topic modeling, which is a class of computational linguistic methods well suited for representing our “semantic neighborhoods” or “models of distributional semantics” (Lupyan and Winter, 2018). Unlike traditional topic models such as Latent Dirichlet Allocation – which are insensitive to multiple topics contained within a single document (topical prevalence) and difference in word choice within a topic across multiple documents (topical content), and rely entirely on the text as opposed to incorporating corpus metadata (Lebryk, 2021) – STMs enable social scientists working with text data to test hypotheses about how topics within this text correlate with non-text variables.

Observations generated in our analysis of the social network – in particular, the Louvain cluster each Twitter user is assigned to – serve as metadata for the structural topic model representing our corpus. Starting with a .csv file representing Tweets from users in a network, each with a corresponding Louvain cluster, the STM package in R allows for the loading, processing, and analysis of topics within this corpus. The first step is inputting the .csv as an R “data frame” – structurally analogous to the similarly-titled Pandas “DataFrame” and thus also easily transformable to and from the .csv format – into our R environment using the read_csv() function built into R. Next we must use a number of functions included in the STM package to prepare and transform the data. The textProcessor() function inputs a data frame (such as the one generated from our .csv of documents and the Louvain cluster metadata for the authors of each document) and outputs another data frame with the necessary inputs for estimating a structured topic model. Most importantly in this step, the textProcessor() function structures the corpus in terms of the three key components of an STM corpus representation:

1. *documents* – a "list containing word indices and their associated counts."
2. *vocab* – a "character vector containing the words associated with the word indices."
3. *metadata* – a "matrix containing document covariates" (Roberts et al., 2019).

After the original text data frame has been processed in this way, it can be further prepared to serve as a basis for STM estimation using the prepDocuments() package also built into the STM library. This package ensures the proper formatting of the data, represented as distinct vectors of an R data frame for each of the three key components of the STM described above. prepDocuments() can also remove the lowest-occurrence words in the corpus, and will properly reindex the resulting R data frame to ensure that any deletions do not result in missing or misaligned connections between documents and metadata. At this point, the resulting data frame is an appropriate input for the stm() function, which estimates the structural topic model for our corpus. A number of functions are built into the STM library for the analysis and visualization of the structural topic model that is generated.

4.6 Procedure

(a) Data Collection

The present study began by identifying a list of key dates over the course of the JCPOA negotiations, as outlined in Table 1. Midnight UTC on these dates was used as the “start_time” parameter for the .search_all_tweets() Tweepy method, with 11:59:59PM UTC seven days later used as the “end_time” parameter. Next, a list of Twitter accounts was identified representing a number of U.S. news media outlets from a wide range of political alignments. A Full-Archive Search was performed in Tweepy by looping over the .search_all_tweet() methods for each of the accounts in our list, and embedding this loop through each account within a loop through all the start dates. The “query” parameter of the method allowed us to collect all Tweets from our target accounts within one week of critical moments in the Iran Deal discourse which mentioned “Iran” and were not Retweets. By looping through all accounts for all dates, we can identify a set of base Tweet objects around which a broader discourse might be identified.

Table 1: Dates of Interest in the JCPOA Discourse

Date	Event
24 November 2013	Finalization of JPA
20 January 2014	JPA goes into effect
2 April 2015	Framework agreement for JCPOA
14 July 2015	JCPOA finalized
20 July 2015	UNSC approves JCPOA resolution, 60-day U.S. Congressional review period begins

17 September 2015	U.S. Congressional review period ends
16 January 2016	IAEA issues certification of Iranian compliance, sanctions on Iran lifted
8 May 2018	Sanctions on Iran reimposed, U.S. withdraws from JCPOA
1 July 2019	Iran informs IAEA it has exceeded enriched uranium stockpile limits imposed by JCPOA
8 September 2019	Iran informs IAEA that it is no longer abiding by R&D limits but continues to allow inspectors

Though the affordances of the Tweet object offer a number of potential routes for reconstructing online communicative dynamics, including analysis of Quote Tweets and Retweets, in the present study we explored networks of accounts replying to Tweets from the media outlets collected in our first round of searching the Twitter archive. To do this, we targeted the “conversation_id” field, which is present for every reply Tweet and identical to the “id” field of the Tweet being replied to. Once again using the .search_all_tweets() method, again excluding Retweets, we identified every Tweet within a week of our start dates which had our original Tweet IDs as its “conversation_id”, indicating that it is a reply to one of the original set of Tweets we collected from media outlets.

After retrieving these Tweets and inputting them into Pandas DataFrames (one for original Tweets from media outlets and one for replies to those Tweets), our discursive core consisted of 1,685 Tweets from 14 users, while our target of investigation – the replies to this discursive core, the users authoring these replies, and their respective positions within the overall social structure of the discourse – consisted of 49,396 Tweets from 5,349 users. Due to limits imposed by computational limits and time constraints (as discussed in Section 4.7), the subsequent analysis was performed using a 380 user subset of this list.

For each of these 380 users, we looped over the .get_users_following() method to compile their entire list of followed accounts, represented as first a DataFrame (i.e. Fig. 3) then a NetworkX edgelist, with the original user as the source and each account they follow as a target. The set of unique entities across source and target nodes is our total node list, while our edge list is each unique connection between a source and a target. For a subset of this node list, as detailed further in Section 5.2, we then performed a final round of Tweet collection to assemble our corpus using the same method used to acquire our discursive core of Tweets from news outlets. The initial network graph consists of 411,048 nodes and 535,911 edges – an incredibly sparse network graph, as is to be expected given the disproportionate ratio of targets per source.

(b) Data Processing

One common technique for identifying structure in sparse networks is *k-core decomposition* (Alvarez-Hamelin et al., 2005), which prunes the network of any nodes connected to fewer than *k* edges. After using NetworkX to perform a k-core decomposition ($k = 5$) on our graph, the pruned graph consisted of 7,884 nodes and 81,128 edges. These pruned node and edge lists were converted to .csv files and used as inputs to generate a graph in Gephi, where further analysis and visualization occurred – the results of which will be discussed in Section 5. The ForceAtlas2 clustering algorithm

(Jacomy et al., 2014) was applied to the graph, using a force-directed layout simulating a physical system wherein “nodes repulse each other like charged particles, while edges attract their nodes” to help translate the statistical topology of a network graph into a spatial perspective. Applying ForceAtlas2 to the subset of our overall network being analyzed here produced Fig.5 below.

The dynamic adjustment and re-visualization of the graph enabled in the Gephi “Overview” tab allows for the use of graph visualization as an exploratory tool, helping researchers understand the underlying structure of their data. In our case, further filtering of nodes by degree reduced the overall crowdedness of the image, resulting in a sparser and technically less informative network which nonetheless helped establish the overall shape of our network. Fig. 6 shows a graph with a minimum degree of 11, resulting in 2,368 nodes (30.04% of the total in the original graph) and 45,439 edges (56.01% of the original total). Fig. 7 shows a further reduced graph with a minimum degree of 25, resulting in 825 nodes (10.46%) and 22,399 edges (27.61%).

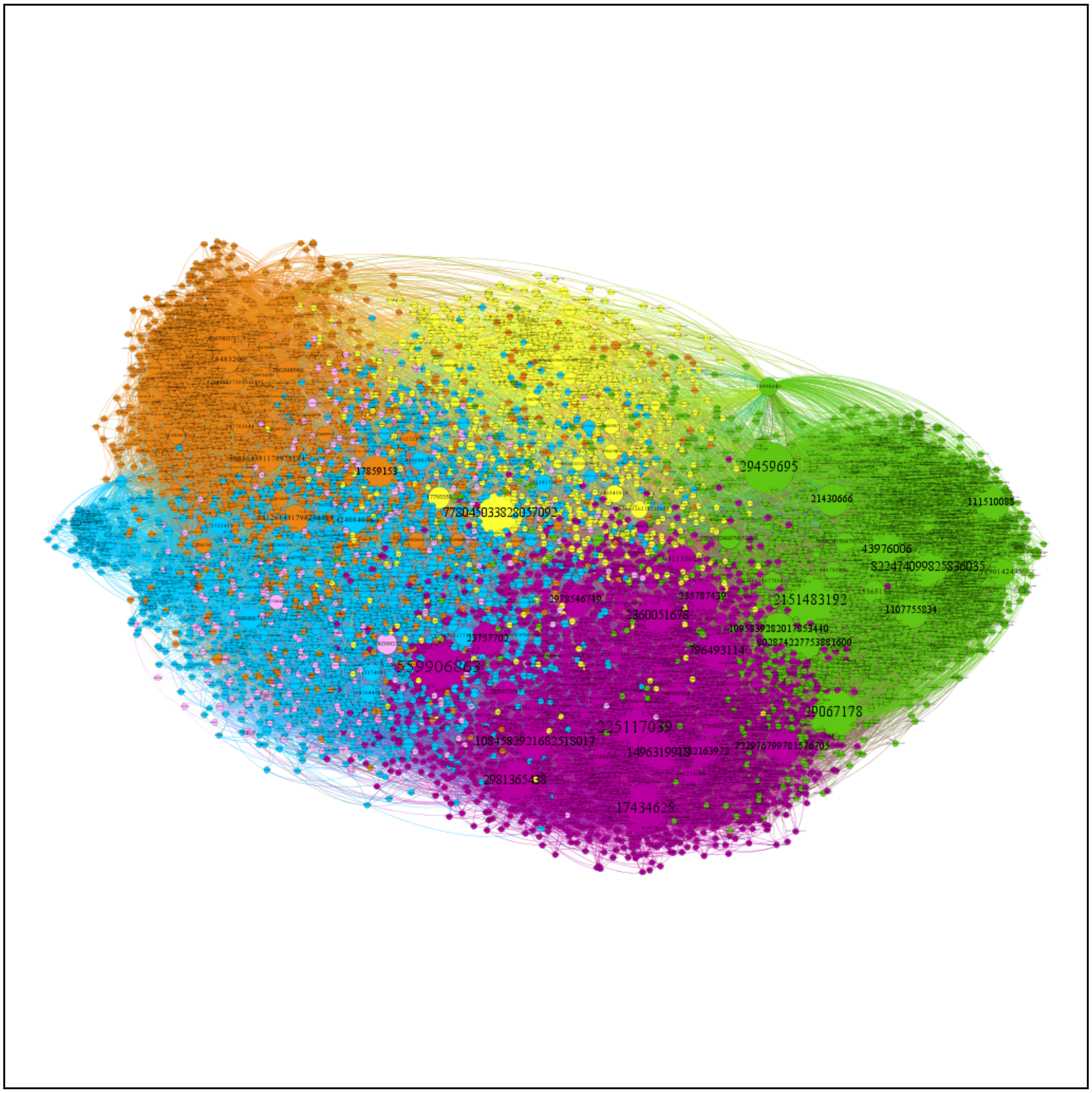

FIG. 5. A network graph generated using ForceAtlas2 in Gephi representing a *k = 5* k-core decomposition of the combined following networks for our 380-user subset. (7,884 nodes and 81,128 edges)

FIG. 6. A network graph generated using ForceAtlas2 in Gephi representing a *k = 11* k-core decomposition of the combined following networks for our 380-user subset. (2,368 nodes and 45,439 edges)

FIG. 7. A network graph generated using ForceAtlas2 in Gephi representing a *k = 25* k-core decomposition of the combined following networks for our 380-user subset. (825 nodes and 22,399 edges)

5. Results

5.1. Gephi Graph Analytics

By partitioning our graph according to the Louvain modularity clusters imported from NetworkX, we were able to generate the overall network topology visualized above as a spatial representation of how the clusters relate to each other. By applying the same ForceAtlas2 algorithm to each partitioned cluster, we were able to better represent the topology of each cluster relative to that of the other clusters.

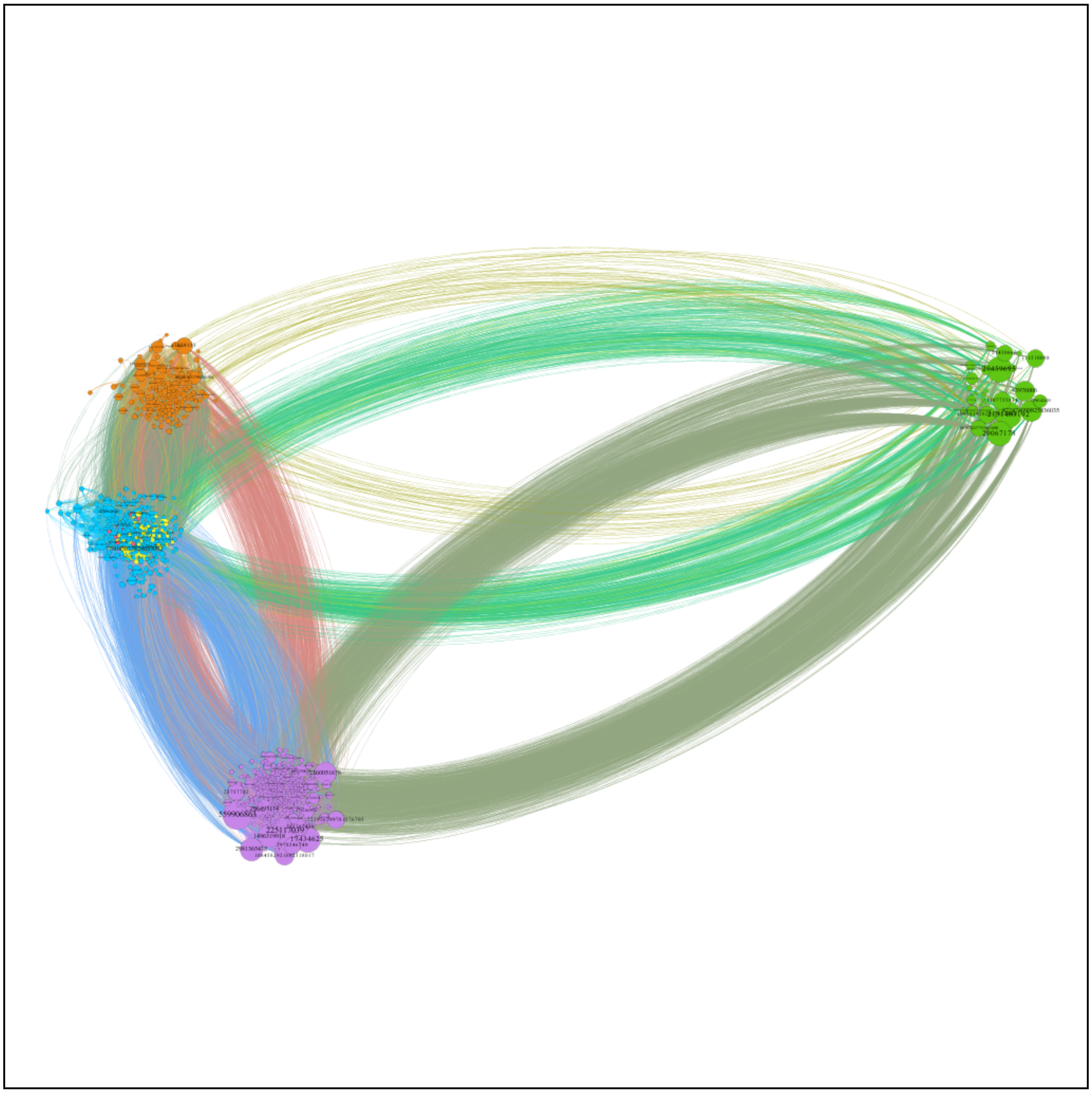

FIG. 8. A network graph generated using ForceAtlas2 in Gephi applied twice, first to a *k = 5* k-core decomposition of the combined following networks for our 380-user subset, then to each of the individual Louvain community clusters represented in that subset. (7,884 nodes and 81,128 edges)

Comparing these internal topologies, we can examine their relative rates of inter-connection (to other nodes within their own cluster) and intra-connection (to nodes in other clusters). For our largest cluster, Group #4, with 28.23% (n = 2,226) of the nodes in the $k = 5$ k-core decomposition network, the internal connections of the group represented 22.31% (n = 18,100) of the overall edges in the entire graph. This demonstrates a high degree of internal cohesion in Group 4, as opposed to Group 3, which is the second most populous Louvain cluster in our graph, and has only 12.37% of our overall edges (n = 10,032) despite including 23.45% of its nodes (n = 1,849). For all our six clusters combined, inter-edges connecting clusters to themselves represented 62.13% of the edges in our overall graph. This indicates a high level of modularity and the presence of discrete clusters. Turning now to intra-edges connecting nodes in one cluster to nodes in another, we can explore which clusters (if any) are looser in their affiliations and interact more extensively with other clusters. Of the overall 37.87% (n = 30,727) of our edges which represent connections from one cluster to another, a majority of 24.25% (n = 19,674) are connections either to or from Group 4. Fig. 9 and Fig. 10 below represent the inter-edge and intra-edge subsets of the $k = 25$ k-core decomposition graph.

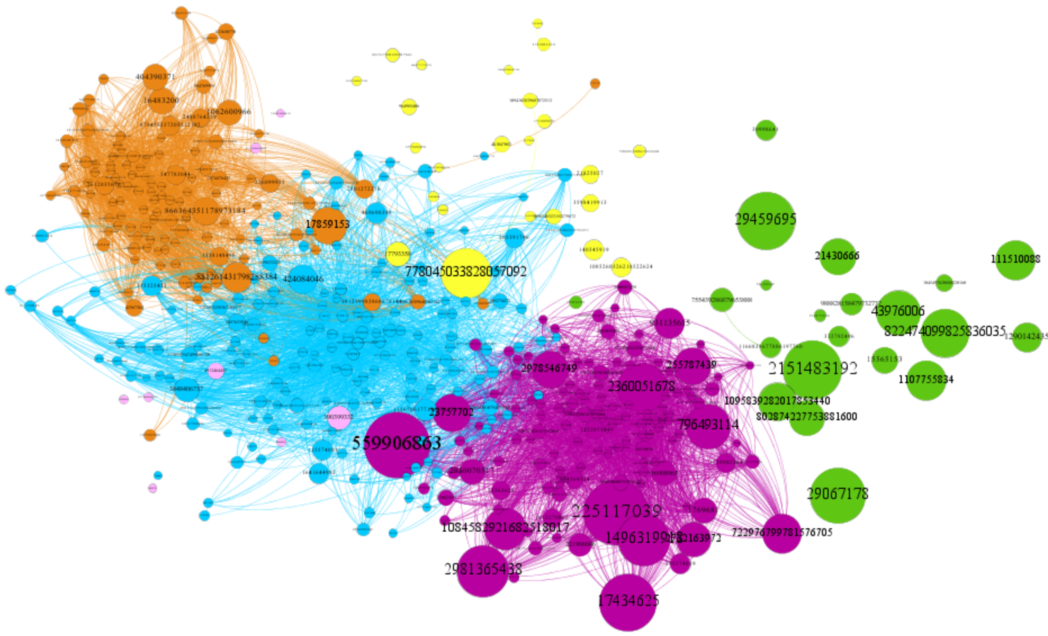

FIG. 9. A network graph generated using ForceAtlas2 in Gephi representating a *k = 25* k-core decomposition of the combined following networks for our 380-user subset, filtered to only include edges connecting nodes within the same cluster. (825 nodes and 50,401 edges)

FIG. 10. A network graph generated using ForceAtlas2 in Gephi representing a *k = 25* k-core decomposition of the combined following networks for our 380-user subset, filtered to only include edges connecting nodes in different clusters. (825 nodes and 30,727 edges)

5.2. Structural Topic Modeling

The textual portion of our analysis was handled in the STM library for R. For a subset of our Twitter user network, a final corpus was generated to associate text content with the Louvain cluster of the user who authored the Tweet. This corpus consisted of 2,393 Tweets from 204 unique users. After the initial steps of STM processing – using the textProcessor() and prepDocuments() functions in the STM library – the elimination of stop words and Tweets which did not contain sufficient data for analysis, this corpus was comprised of 2,387 Tweets.

The function of the structural topic model is to generate a word vector representing the overall "dictionary" for the corpus. Counts of token mapped onto this dictionary then serves as the basis for constructing vectors which are then used to correlate the content and prevalence of topics in the model. Using the summarize() function built into R, we can identify key words for each topic identified by the model (e.g. Fig. 11), key words for each Louvain cluster, as well as key words for each topic sorted by cluster.

Topic 1: iran, deal, nuclear, new, will, say, talk, state, negoti, foreign, reviv, agreement, sanction, tehran, minist, biden, world, vienna, democrat, power
Topic 2: iran, irgc, terror, terrorist, guard, revolutionari, biden, iranâ€™, nazanin, zaghari-ratcliff, year, american, remov, delist, state, list, releas, organ, foreign, design
Topic 3: iran, updat, israel, live, world, two, toialert, gas, say, today, news, will, blinken, russian, women, cup, nuke, jasonmbrodski, govern, get
Topic 4: iran, iranian, attack, regim, amp, houthi, missil, iran-back, iraq, israel, saudi, right, offici, say, isra, support, women, report, militia, prison
Topic 5: iran, biden, russia, sanction, putin, ukrain, oil, regim, amp, administr, weakerirand, will, war, appeas, weapon, venezuela, china, billion, saudi, get

FIG. 11. A table indicating key words for each of our STM-generated topics.

While some topics can be readily inferred from just this list of keywords – e.g. Topic 1 seems to be about the negotiation of the deal itself, while Topic 5 seems to be about how the deal fits into a broader geopolitical context – further investigation of the generated topics is required for a more comprehensive understanding of their contents and boundaries.

Differences in word choice between different Louvain groups discussing the same topic – the major object of investigation in this pilot study – remain elusive enough that it would be premature at this moment to assess the role of social clustering in driving divergent conceptualization in the JCPOA discourse. Nonetheless, cursory inspection of the topical content covariate summary indicates that, at least for Topic 5, substantial differences with distinct political valences can be detected between different subgroups of Twitter users.

```
Topic 5, Group 0: reinstat, bolivia, uzbekistan, dprk, congo, gabon, lao
Topic 5, Group 1: confus, didnâ€™t, dni, gsmittysmith, johnratcliff, oâ€™brien, removingâ€¦
Topic 5, Group 2: â€¢, anasalhajji, al-sadhan, ding, felix, gulshan, jiaxi
Topic 5, Group 3: senjohnbarrasso, coal, elitist, puppet, regul, repboebert, roughneck
Topic 5, Group 4: appeas, get, billion, reward, cash, thank, influenc
Topic 5, Group 5: houthi, diminish, terrorismâ€™, banker, emptywheel, jackfro, â€˜central
```

FIG. 12. Topic 5 from the structural topic model, separated into a distinct list of keywords for each one of our Louvain community detection algorithm groups.

Though it remains unclear exactly what users in these subgroups are discussing, it’s clear that Group 0 and Group 5 are focused on different sets of international actors. Groups 1, 2, and 3 seem to be discussing different sets of individuals – raising questions about text processing that will be addressed in the next section. Finally, Group 4 is primarily characterized by the language of commercial exchange and moral evaluation, very likely in the context of sanctions against Iran.

6. Discussion and Conclusion

What bearing do these results have for our understanding of frame semantic entrenchment within a discourse? The distributional semantics of key topics in this discourse indicates differential patterns of linguistic entrenchment among social clusters for ostensibly shared topics. If we understand linguistic entrenchment, borrowing from Lupyan and Winter (2018), as actively carving conceptual structure out of the world, we would expect that these patterns in distributional semantics centered around shared abstract concepts – e.g. Iranian nuclear diplomacy – would indicate diverging understandings of those concepts. Indeed, without the discursive environments in which these patterns emerge, such understandings could hardly develop.

As a potential example of the discursive face of this conceptual tumult, we can examine social clusters within Topic 5 and how the most salient aspects of this topic vary among these groups. The different groups of international actors profiled in Groups 0 and 5 suggest very different sets of concerns surrounding the impact of the Iranian nuclear deal on global politics. For Group 0, the focus is on countries around the world, such as Gabon and Congo (in Africa) and, most notably, key U.S. rivals Bolivia (in South America) and North Korea ("dprk" in East Asia). This might suggest that for Group 0, the JCPOA primarily evokes the precedents it sets for the behavior of other states, and that this interpretation is cemented through repeated discussion of the deal in the context of international norms. Group 5 is more focused on actors and actions involved in regional politics in and around Iran – the Iranian-backed Houthis fighting Saudi Arabia and "bankers," as well as "terrorism." For this group, the JCPOA becomes conceptually entrenched as a tool of military diplomacy, a limit to Iranian ability to influence regional affairs. Perhaps notably, this divide maps onto liberal and realist (e.g. Mowle 2003) schools of international relations.

While anything more than a tentative exploration of these divisions – including an analysis of their change over time – is not possible without further investigation, we might expect that the conceptual structures entrenched in any one social cluster might resist change via integration with competing conceptual structures, following from Bybee and Thompson's (1997) Conserving Effect. Given the role of frequency in reifying morphological structure, and given the basic psychological processes underlying both linguistic and conceptual entrenchment, we can expect that frequency of conceptual structure also leads to similar resistance to change as is seen in "slept" and "kept" but not "creeped" or "weeped." Just like their lexical counterparts, entrenched conceptual forms may become less likely to change through frequency and repetition. To the extent that this conceptual structure is shaped by the distributional semantics of discourse in a social cluster, these differently reified conceptual forms will correlate with differences in co-occurrence of specific lexical units between clusters, which we have measured here through structural topic modeling.

Perhaps the most significant improvement to the current investigation could be achieved with the inclusion of more time and computational resources. For certain parts of the data collection process – particularly the use of the .search_all_tweets() and .get_users_following() methods – rate limits imposed by the Twitter API required substantial truncation of data collection at times and the use of subsets of larger datasets. The use of subsets also became necessary when constructing networks from one-degree ego networks centered around our corpus of replies to news outlets, due not to rate limits but rather to computational limits imposed both by hardware and by software. For the generation of more populous networks, high-performance computing and more robust tools for network analysis than NetworkX may be required.

Another area for potential improvement as this method is iterated upon – perhaps applied to different discourses where divergent conceptualization might also be evident – is in the generation of the initial Tweet corpus. Rather than relying on my intuition as an experimenter, a more independent first step would certainly enable more reliable user clustering and subsequent analysis. A useful tool for assembling this more independent initial corpus could be the “Stream” functionality of the Twitter API, which will collect a sample of all Tweets matching a particular query for the entire time the API user remains connected to the Stream. For queries with low response rates, this method can serve to acquire a constant input of linguistic and socio-pragmatic information regarding the online contours of a particular discourse.

Finally, the introduction of new sources of linguistic data and metadata to incorporate into the process would enhance our ability to target specific linguistic structures. Prominent corpus linguistic tools include the COBUILD project (Sinclair, 2004) and the FrameNet project (Baker, et al., 1998). The FrameNet project collects instances of lexical frames and identifies common lexical units for each frame. Whereas COBUILD could serve as a useful reference corpus to compare with the patterns of linguistic entrenchment found in our target corpus, FrameNet could be introduced as part of the data processing pipeline to categorize documents in the corpus based on the evocation of a given frame.

As a two-pronged investigation into both a linguistic question about conceptual entrenchment, as well as a methodological question about corpus linguistics using Twitter data, this project was highly exploratory and suggests many avenues for further development and application of these theories and tools. In continuing to explore linguistic markers of divergent conceptualization, the role of entrenchment ensures that this particular study of concepts remains grounded in the empiricism and usage-first perspective of construction grammar and corpus linguistics.

Works Cited

Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). k-core decomposition: A tool for the visualization of large scale networks. *ArXiv:Cs/0504107*. http://arxiv.org/abs/cs/0504107

Austin, J. L. (1962). *How to Do Things with Words*. Clarendon Press.

Ayer, A. J. (1940). *The Foundations Of Empirical Knowledge*. London, England: Macmillan.

Baker, C. F., Fillmore, C. J., & John B. Lowe. (1998). The Berkeley FrameNet project. *COLING-ACL '98: Proceedings of the Conference*. COLING-ACL '98: Proceedings of the Conference, Montreal, Canada.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361–362.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bybee, J. L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. John Benjamins Publishing.

Bybee, J., & Thompson, S. (1997). Three Frequency Effects in Syntax. *Annual Meeting of the Berkeley Linguistics Society*, *23*(1), 378–388. https://doi.org/10.3765/bls.v23i1.1293

Chomsky, N. (1965). *Aspects of the theory of syntax*. M.I.T. Press.

Coulson, S. (2006). Conceptual Blending in Thought, Rhetoric, and Ideology. In G. Kristiansen, M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (Vol. 1–xi, 499, pp. 187–208). Mouton de Gruyter; MLA International Bibliography. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=mzh&AN=2007930838&site=eds-live&custid=s8481523

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of Pragmatics*, *37*(10), 1510–1536. https://doi.org/10.1016/j.pragma.2004.09.010

Crane, T., & French, C. (2021). The Problem of Perception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/perception-problem/

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PLoS ONE*, *9*(11), e113114. https://doi.org/10.1371/journal.pone.0113114

Eisenstein, S. (1949). *A Dialectic Approach to Film Form*. http://antigo.casaruibarbosa.gov.br/arquivos/file/A_Dialectic_Approach_to%20_Film_Form_SergeiEisenstein.pdf

Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. L. Erlbaum.

Fauconnier, G. (1997). *Mappings in thought and language* (pp. ix, 205). Cambridge University Press. https://doi.org/10.1017/CBO9781139174220

Fauconnier, G., & Turner, M. B. (2001). *Conceptual Integration Networks* (SSRN Scholarly Paper ID 1292966). Social Science Research Network. https://papers.ssrn.com/abstract=1292966

Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the Morning Calm*. Hanshin Publishing Co.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, *64*(3), 501–538. https://doi.org/10.2307/414531

Firth, R. (1964). Austin and the Argument from Illusion. *The Philosophical Review*, *73*(3), 372–382. https://doi.org/10.2307/2183663

Full-archive search—Metadata and filtering timeline. (n.d.). Retrieved April 15, 2022, from https://developer.twitter.com/en/docs/twitter-api/enterprise/search-api/guides/changelog

Gaumont, N., Panahi, M., & Chavalarias, D. (2018). Reconstruction of the socio-semantic dynamics of political activist Twitter networks-Method and application to the 2017 French presidential election. *PloS One*, *13*(9), e0201879. https://doi.org/10.1371/journal.pone.0201879

Geeraerts, D., Grondelaers, S., & Bakema, P. (2012). The Structure of Lexical Variation: Meaning, Naming, and Context. In *The Structure of Lexical Variation*. De Gruyter Mouton. https://doi.org/10.1515/9783110873061

Glenney, B., & Silva, J. F. (2019). *The Senses and the History of Philosophy*. Routledge.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Gries, S. T. (2012). *Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges* (pp. 41–63). Brill. https://doi.org/10.1163/9789401207713_006

Gries, S. T. (2017). Corpus Linguistics, Cognitive Linguistics, and Psycholinguistics: On their Combination and Fit. In *Ten Lectures on Quantitative Approaches in Cognitive Linguistics* (pp. 1–22). Brill. https://doi.org/10.1163/9789004336223_002

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference* (pp. 11–15).

Haiman, J. (1994). *Ritualization and the development of language*. https://doi.org/10.1075/cilt.109.07hai

Hooper, J. B. (1976). *An Introduction to Natural Generative Phonology*. Academic Press.

Inc, G. (2008, February 27). *Iran*. Gallup.Com. https://news.gallup.com/poll/116236/Iran.aspx

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, *9*(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Kuleshov, L. (1935). *Principles of Montage*. https://www.sas.upenn.edu/~cavitch/pdf-library/Kuleshov_Principles%20of%20Montage.pdf

Lacey Okonski, & Luciane Corrêa Ferreira. (2019). Gonna be on my fucking period in boomtown, souuuunndd thanks Mother Nature: Using Twitter to Find Multimodal Creativity and Embodied Instant Metaphors. *Signo*, *44*(79), 122–134. Directory of Open Access Journals. https://doi.org/10.17058/signo.v44i79.12851

Lebryk, T. (2021, April 18). *Introduction to the Structural Topic Model (STM)*. Medium. https://towardsdatascience.com/introduction-to-the-structural-topic-model-stm-34ec4bd5383

Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, *8*(1), 339–359. https://doi.org/10.1007/bf00258436

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170137. https://doi.org/10.1098/rstb.2017.0137

Markle, F. (1964). A Talk with Hitchcock. In *Telescope*.

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Most Americans think Iran wants nuclear weapons, poll says—CNN.com. (n.d.). Retrieved May 11, 2021, from http://edition.cnn.com/2009/POLITICS/10/20/us.iran.poll/

Mowle, T. S. (2003). Worldviews in Foreign Policy: Realism, Liberalism, and External Conflict. Political Psychology, 24(3), 561–592.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

R Core Team. (2021). *R: The R Project for Statistical Computing*. https://www.r-project.org/

Reyes-Menendez, A., Saura, J. R., & Thomas, S. B. (2020). Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *International Journal of Information Management*, *54*, 102129. https://doi.org/10.1016/j.ijinfomgt.2020.102129

Reynolds, S. L. (2000). The Argument from Illusion. *Noûs*, *34*(4), 604–621.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). **stm**: An *R* Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2). https://doi.org/10.18637/jss.v091.i02

Römer, U., O'Donnell, M., & Ellis, N. C. (2015). *Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge.* John Benjamins. http://deepblue.lib.umich.edu/handle/2027.42/139835

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization*. Lawrence Elbaum Associates.

Rosenbach, E., & Peritz, A. (2009, July). *Confrontation or Collaboration? Congress and the Intelligence Community*. Belfer Center for Science and International Affairs. https://www.belfercenter.org/publication/confrontation-or-collaboration-congress-and-intelligence-community

Šćepanović, S., Mishkovski, I., Gonçalves, B., Nguyen, T. H., & Hui, P. (2017). Semantic homophily in online communication: Evidence from Twitter. *Online Social Networks and Media*, *2*, 1–18. https://doi.org/10.1016/j.osnem.2017.06.001

Schmid, H.-J. (2010, June 9). *Entrenchment, Salience, and Basic Levels*. The Oxford Handbook of Cognitive Linguistics. https://doi.org/10.1093/oxfordhb/9780199738632.013.0005

Schmid, H.-J. (2015). A blueprint of the Entrenchment-and- Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association*, *3*(1). https://doi.org/10.1515/gcla-2015-0002

Sheffield, M. (2019, June 24). Poll: 24 percent of voters want military action against Iran [Text]. *The Hill*. https://thehill.com/hilltv/what-americas-thinking/450050-poll-only-24-percent-of-americans-want-us-to-take-military/

Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. 1–212. https://doi.org/10.4324/9780203594070

Stubbs, M. (2007). Memorial Article: John Sinclair (1933-2007): The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics - APPL LINGUIST*, *30*, 115–137. https://doi.org/10.1093/applin/amn052

Sue Atkins, Fillmore, C. J., & Christopher R. Johnson. (2003). Lexicographic relevance: Selecting information from corpus evidence. *International Journal of Lexicography*, *16.3*.

Telles-Correia, D., Moreira, A. L., & Gonçalves, J. S. (2015). Hallucinations and related concepts—Their conceptual background. *Frontiers in Psychology*, *6*, 991. https://doi.org/10.3389/fpsyg.2015.00991

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00725

Tobin, V. (2018). *Elements of surprise: Our mental limits and the satisfactions of plot* (p. 332). Harvard University Press. https://doi.org/10.4159/9780674919570

Tobin, V. (2020). Experimental investigations of irony as a viewpoint phenomenon. In *Experimental investigations of irony as a viewpoint phenomenon* (pp. 236–255). De Gruyter Mouton. https://doi.org/10.1515/9783110652246-011

Tobin, V. (2021). Where irony goes: Routinization and the collapse of viewpoint configurations. *Chinese Semiotic Studies*, *17*, 199–227. https://doi.org/10.1515/css-2021-0011

Tobin, V., & Israel, M. (2012). Irony as a viewpoint phenomenon. *Viewpoint in Language: A Multimodal Perspective*, 25–46. https://doi.org/10.1017/CBO9781139084727.004

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition* (pp. viii, 388). Harvard University Press.

Turner, M. B. (2006). *Frame Blending* (SSRN Scholarly Paper ID 1321302). Social Science Research Network. https://papers.ssrn.com/abstract=1321302

Wasow, T. (2015). Ambiguity avoidance is overrated. In *Ambiguity: Language and Communication*. Walter de Gruyter GmbH; Scopus®. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edselc&AN=edselc.2-52.0-84960218662&site=eds-live&custid=s8481523

Wasow, T., Perfors, A., & Beaver, D. (2003). *The Puzzle of Ambiguity*.

Zenko, M. (2015, April 2). *Iran Has Been Two Years Away From a Nuclear Bomb Since the 1980s*. The Atlantic. https://www.theatlantic.com/international/archive/2015/04/iran-has-been-two-years-away-from-a-nuclear-bomb-since-the-1980s/389333/

Zottola, A. (2020). Corpus Linguistics and Digital Humanities. Intersecting Paths. A Case Study from Twitter. *América Crítica*, *4*(2), 131–141. https://doi.org/10.13125/americacritica/4521