A Game-Theoretic Model of Metaphorical Bargaining

Beata Beigman Klebanov
Kellogg School of Management
Northwestern University
`beata@northwestern.edu`

Eyal Beigman
Washington University in St. Louis
`beigman@wustl.edu`

Abstract

We present a game-theoretic model of bargaining over a metaphor in the context of political communication, find its equilibrium, and use it to rationalize observed linguistic behavior. We argue that game theory is well suited for modeling discourse as a dynamic resulting from a number of conflicting pressures, and suggest applications of interest to computational linguists.

1 Introduction

A 13 Dec 1992 article in *The Times* starts thus:

> The European train chugged out of the station last night; for most of the day it looked as if it might be stalled there for some time. It managed to pull away at around 10:30 pm only after the Spanish prime minister, Felipe Gonzalez, forced the passengers in the first class carriages into a last minute whip round to sweeten the trip for the European Community's poor four: Spain, Portugal, Greece and Ireland.
>
> The fat controller, Helmut Kohl, beamed with satisfaction as the deal was done. The elegantly-suited Francois Mitterrand was equally satisfied. But nobody was as pleased as John Major, stationmaster for the UK presidency, for whom the agreement marked a scarce high point in a battered premiership.
>
> The departure had actually been delayed by seven months by Danes on the line. Just when that problem was solved, there was the voluble outbreak, orchestrated by Spain, from the poor four passengers demanding that they should travel free and be given spending money, too.
>
> The coupling of the carriages may not be reliably secure but the pan-European express is in motion. That few seem to agree the destination suggests that future arguments are inevitable at every set of points. Next stop: Copenhagen.

Apart from an entertaining read, the extended metaphor provides an elaborate conceptual correspondence between a familiar domain of train journeys and the unfolding process of European integration. Carriages are likened to nation states; passengers to their peoples; treaties to stations; politicians to responsible rail company employees.

In a compact form, the metaphor gives expression to both the small and the large scale of the process. It provides for the recent history: Denmark's failure to ratify the 1992 Maastricht treaty until opt-outs were negotiated later that year is compared to dissenters sabotaging the journey by laying on the tracks (*Danes on the line*); negotiations over the Cohesion Fund that would provide less developed regions with financial aid to help them comply with convergence criteria are likened to second class carriages with poor passengers for whom the journey had to be subsidized. At a more general level, the European integration is a purposeful movement towards some destination according to a worked out plan, getting safely through negotiation and implementation from one treaty to another, as a train moving on its rails through subsequent stations, with each nation being separate yet tied with everyone else. Numerous inferences regarding speed, timetables, stations, passengers, different classes of tickets, temporary obstacles on the tracks, and so on can be made by the reader based on the knowledge of train journeys, giving him or her a feeling of an enhanced understanding[1] of the highly complex process of European integration.

So apt was the metaphor that political fights were waged over its details (Musolff, 2000). Worries about destination were given an eloquent expression by Margaret Thatcher (*Sunday Times*, 20 Sept 1992):

> She warned EC leaders to stop their endless round of summits and take notice of their own people. "There is a fear that the European train will thunder forward, laden with its customary cargo of gravy, towards a destination neither wished for nor understood by electorates. But the train can be stopped," she said.

---

[1]More on enhanced understanding in sections 3.2 and 4.2.

The metaphor proved flexible enough for further elaboration. John Major, a Conservative PM of Britain, spoke on June 1st, 1994 about his vision of the decision making at the EU level, saying that he had never believed that Europe must act as one on every issue, and advocating "a sensible new approach, varying when it needs to, multi-track, multi-speed, multi-layered." He attempted to turn a largely negative Conservative take on the European train (see Thatcher above) into a tenable positive vision — each nation-carriage is now presumably a rather autonomous entity, waiting on a side track for the right locomotive, in a huge yet smoothly operating railroad system.

Major's political opponents offered their counter-frames. In both cases, the imagery of a large transportation system was taken up, yet turned around to suggest that "multi, for everyone" amounts to Britain being in "the slow lane," and a different image was suggested that makes the negative evaluation of Britain's opt-outs more poignant — a football metaphor, where relegation to the second division is a sign of a weak performance, and a school metaphor, where Britain is portrayed as an under-achiever:

John Cunningham, Labour He has admitted that his Government would let Britain fall behind in Europe. He is apparently willing to offer voluntary relegation to the second division in Europe, and he isn't even prepared to put up a fight. I believe that in any two-speed Europe, Britain must be up with those in the fast lane. Clearly Mr Major does not.

Paddy Ashdown, Liberal Democrat Are you really saying that the best that Britain can hope for under your leadership is ... the slow lane of a two-speed Europe? Most people in this country will want to aim higher, and will reject your view of a 'drop-out' Britain.

The pro-European camp rallied around the "Britain in the slow lane" version as a critical stance towards the government's European policy. Of the alternative metaphors, the school metaphor has some traction in the Euro discourse, where the European (mainly German) financial officers are compared to school authorities, and governments struggling to meet the strict convergence criteria to enter the Euro are compared to pupils that barely make the grade with Britain as a 'drop-out' who gave up even trying (Musolff, 2000).

The fact that European policy is being communicated and negotiated via a metaphor is not surprising; after all, "there is always someone willing to help us think by providing us with a metaphor

that accords with HIS views."[2] From the point of view of the dynamics of political discourse, the puzzle is rather the apparent tendency of politicians to be compelled by the *rival's* metaphorical framework. Thatcher tries to turn the train metaphor used by the pro-EU camp around. Yet, assuming metaphors are matters of choice, why should Thatcher feel constrained by her rival's choice, why doesn't she ignore it and merely suggest a new metaphor of her own design? As the evidence above suggests, this is not Thatcher's idiosyncrasy, as Major and his rivals acted similarly. Can this dynamic be explained?

In this article, we use the explanatory framework of game theory, seeking to rationalize the observed behavior by designing a game that would produce, at equilibrium, the observed dynamics. Specifically, we formalize the notion that the price of "locking" the public into a metaphorical frame of reference is that a politician is coerced into staying within the metaphor as well, even if he or she is at the receiving end of a rival's rhetorical move.

Since the use of game theory is not common in computational linguistics, we first explain its main attributes, justify our decision to make use of it, and draw connections to research questions that can benefit from its application (section 2). Next, we design the game of bargaining over a metaphor, and find its equilibrium (section 3), followed by a discussion (section 4).

2 Game-Theoretic models

The basic construct is that of a game, that is, a model of participants in an interaction (called "players"), their goals (or "utilities") and allowable moves. Different moves yield different utilities for a player; it is assumed that each player would pick a strategy that maximizes her utility. The observable is the actual sequence of moves; importantly, these are assumed to be the optimal outcome (an equilibrium) of the relevant game. A popular notion of equilibrium is Nash equilibrium (Nash, 1950). For extensive form games (the type employed in this paper), the notion of *subgame perfect equilibirum* is typically used, denoting a Nash equilibrium that would remain such if the players start from any stage of the evolving game (Selten (1975; 1965)).

The task of a game theorist is to reverse-engineer the model for which the observed se-

---

[2]Capitalization in the original, Bolinger (1980, p. 146).

quence of actions is an equilibrium. The resulting model is thereby able to rationalize the observed behavior as a naturally emerging dynamics between agents maximizing certain utility functions. In economics, game-theoretic models are used to explain price change, organization of production, and market failures (Mas-Colell et al., 1995; von Neumann and Morgenstern, 1944); in biology — the operation of natural selection processes (Axelrod and Hamilton, 1981; Maynard Smith and Price, 1973); in social sciences — political institutions, collective action, and conflict (Greif, 2006; Schelling, 1997; North, 1990). In recent applications in linguistics, pragmatic phenoma such as implicatures are rendered as an equilibrium outcome of a communication game (Jäger and Ebert, 2008; van Rooij, 2008; Ross, 2007; van Rooij and Schulz, 2004; Parikh, 2001; Glazer and Rubinstein, 2001; Dekker and van Rooy, 2000).

Computing equilibria is simple for some games and quite evolved for others. For example, computing the equilibrium of a zero-sum game is equivalent to LP optimization (Luce and Raiffa, 1957); an equilibrium of general bimatrix games can be found using a pivoting algorithm (von Stengel, 2007; Lemke and Howson, 1964). Interesting connections have been pointed out between game theory and machine learning: Freund and Schapire (1996) present both online learning and boosting as a repeated zero-sum game; Shalev-Shwartz and Singer (2006) show similarly that loss minimization in online learning is akin to an equilibrium path in a repeated game.

While game theoretic models are not much utilized in computational linguistics, they are quite attractive to tackle some of the problems computational linguists are interested in. For example, generation of referring expressions (Paraboni et al., 2007; Gardent et al., 2004; Siddharthan and Copestake, 2004; Dale and Reiter, 1995) can be rendered as a communication game with utility functions that reflect pressures to use shorter expressions while avoiding excessive ambiguity (Clark and Parikh, 2007), with corpora annotated for entity mentions informing the design of a model. Generally, computational linguistics research produces algorithms to detect entities of various kinds, be it topics, named entities, metaphors, moves in a multi-party conversations, or syntactic constructions in large corpora; such primary data can be used to trace developments

not only in chronological terms (Gruhl et al., 2004; Allan, 2002), but in strategic terms, i.e. in terms that reflect agendas of the actors, such as political agendas in legislatures (Quinn et al., 2006) or activist forums (Greene and Resnik, 2009), research agendas in group meetings (Morgan et al., 2001), or social agendas in speed-dates (Jurafsky et al., 2009). Game theoretical models are well suited for modeling dynamics that emerge under multiple, possibly conflicting constraints, as we exemplify in this article.

3   The model

We extend Rubinstein (1982) model of negotiation through offers and counter-offers between two players with a public benefit constraint.

The model consists of (1) two players representing the opposing sides, (2) a set of **frames** $X \subset \mathbb{R}^n$ compact and convex, (3) preference relations described by continuous utility functions $U_1, U_2 : X \to \mathbb{R}^+$, (4) a sequence of frames $\mathcal{X}_0 \subset \mathcal{X}_1 \ldots \subset 2^X$ that can be suggested to the public, and (5) a sequence of public preferences over frames in $\mathcal{X}_t$ for $t = 0, 1, 2, \ldots$ described by a public utility function U_t^p.

The game proceeds as follows. Initially the frame is $F_0 = X$. In odd rounds player 1 appeals to the public with a frame $A_t^1 \in \mathcal{X}_{t|F_t}$, $\mathcal{X}_{t|F_t} = \{A \in \mathcal{X}_t : A \subset F_t\}$, player 2 counters with a frame $A_t^2 \in \mathcal{X}_{t|F_t}$. The public chooses one of the frames based on $U_t^p(A_t^i)$ with ties broken in 1's favor. The accepted frame becomes the current frame for the next round F_{t+1}. In even rounds the parts of players 1 and 2 are reversed.

A finite sequence $F_0, \ldots, F_{t-1}$ gives the *history* of the bargaining process up to t. A *strategy* σ_i of player i is a function specifying for any history $h = \{F_0, \ldots, F_{t-1}\}$ the move player i makes at time t, namely the frame A_t^i she chooses to address the public. A sequence $F_0, F_1, F_2, F_3, \ldots$ describes a path the bargaining process can take, leading to an *outcome* $\cap_{t=0}^{\infty} F_t$. The players' utility for an outcome is given by $U_i = \lim_{t \to \infty} \int_{F_t} U_i(x) d\chi_{F_t}$ for $i = 1, 2$ where χ_{F_t} is a probability measure on F_t. If $\cap_{t=0}^{\infty} F_t = \{x\}$ the utility is the point utility of x otherwise it is the expected utility on the intersection set.

3.1   Player utility

For a given issue under discussion, such as European integration process, we order the possible

states of the world along a single dimension that spans the policy variations proposed by the different players (politicians). Politics of a single issue are routinely modeled as lying on a single dimension.[3] In the British context, various configurations of the unfolding European reality are situated along the line between high degree of integration and complete separatism; Liberal Democrats are the most pro-European party, while United Kingdom Independence Party are at the far-right end of the scale, preferring British withdrawal from the EU. The two major parties, Labour and Conservatives (Tories), prefer intermediate left-leaning and right-leaning positions, respectively. A schematic description is shown in figure 1.

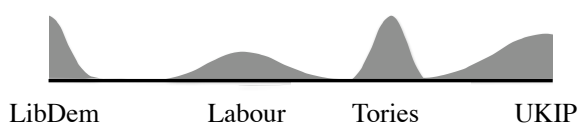LibDem   Labour   Tories   UKIP

Figure 1: Preferences on pro-anti Europe axis.

The utilities of the different players can in this case be described as continuous single-peaked functions over an interval.[4] Thus $X=[0,1]$, and the utility functions $U_i(x)=\phi(||x-v_i||)$ for $v_i \in X$ where ϕ is a monotonically strictly decreasing function and $|| \, ||$ is Euclidean distance.

3.2 Public utility

We note the difference between two types of utilities: The utility of the players is over outcomes, the utility of the public is over sets of outcomes (frames). The latter does not represent a utility the public has for one outcome or another, but rather a utility it has for an enhanced understanding. Thus, the public's utility from a frame is a function of the information content of the proposed frame relative to the current frame, i.e. the relative entropy of the two sets.[5] Formally, if the accepted

frame at time t is F_t then for any Borel set $A \subset F_t$ the public utility for A is $U_t^p(A)=\Pi(\mathrm{Ent}_t(A))$ where $\mathrm{Ent}_t(A)=-\mu_t(A)\log\mu_t(A)$ for a continuous probability measure μ_t on F_t and Π is a continuous, monotone ascending function; for $A \not\subset F_t$, $U_t^p(A)=0$. We take μ_t to be the relative length of the segment $\mu_t(A)=\frac{|A|}{|F_t|}$, hence the entropy maximizing subsegments are of length $\frac{|F_t|}{2}$.

3.3 Game dynamics

At every point in the game, a certain set of the states-of-affairs is being deemed sufficiently probable by the public to require consideration. Suppose that initially any state of affairs within the interval $[0,1]$ is assigned a uniform probability and thus merits public attention. Each in her turn, the players propose to the public to concentrate on a subset of the currently considered states of affairs, arguing that those are the likelier ones to obtain, hence merit further attention. The metaphor used to deliver the proposal describes the newly proposed subset in a way that makes those states-of-affairs that are in it aligned with the metaphor, whereas all other states are left out of the proposed metaphorical frame. As the game proceeds, the public attention is concentrated on successively smaller sets of eventualities, and these are given a more and more detailed metaphoric description, providing the educational gratification of increasingly knowing better and better what is going on. At each step, each player strives to provide maximum public gratification while leading the public to focus on the frame (i.e. subset of states of affairs) that best meets the player's preferences.[6]

Figure 2 sketches the frame negotiation through train metaphor, from some point in time when the general train metaphor got established, through Thatcher's flashing out the issue of excessive speed and unclear direction, Major's multi-track corrective, and reply of his opponents on the left. The final frame has all those states of affairs that fit the extended metaphor – everyone is acting within the same broad system of rules, with Britain and perhaps others sometimes wanting to negotiate special, more gradual procedures, which would leave Britain less tightly integrated into the com-

---

[3]Indeed, Poole and Rosenthal (1997) argue that no more than two dimensions are needed to account for voting patterns on *all* issues in the US Congress.

[4]Single-peakedness is a common assumption in position modeling in political science (Downs, 1957).

[5]The notion that new beliefs are refinements of existing ones is current in contemporary theorizing about formation and change of beliefs, evaluations, and preferences. An update based on the latest available information is consistent with memory-based theories; in our model, in the equilibrium, the current frame contains information about the path-so-far, thus early stages of the bargaining processes are in some sense integrated into the current frame, compatible with the rival, online model of belief formation. See Druckman and Luria (2000) for a review of the relevant literature.

[6]We note that in our model every utterance has an impact on the public for which the player bears the consequences and is therefore a (costly) strategic move in the game. This is different from models of cheap talk such as Aumann (1990), Lewis (1969) where communication is devoid of strategic moves and is used primarily as a coordination device.
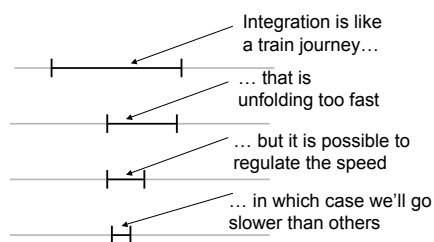
munity than some other European partners.

Figure 2: Bargaining over train metaphor.

3.4 The equilibrium

A pair of strategies (σ_1, σ_2) is a *Nash equilibrium* if there is no deviation strategy σ such that (σ, σ_2) leads to an outcome with higher utility for player 1 than outcome of (σ_1, σ_2) and the same for player 2. A *subgame* are all the possible moves following a history $h=\{F_0, \dots, F_t\}$, in our case it is equivalent to a game with an initial frame F_t and the corresponding utilities. A *sub-strategy* is that part of the original strategy that is a strategy on the subgame. A pair of strategies is a *subgame perfect equilibrium* if, for any subgame, their sub-strategies are a Nash equilibrium.

Theorem 1 *In the frame bargaining game with single-peaked preferences*

1. *There exists a canonical subgame perfect equilibrium path* $F_0, F_1, F_2, \dots$ *such that* $\cap_{t=0}^{\infty} F_t = \{x\}$.

2. *For any subgame perfect equilibrium path* $F_0', F_1', F_2', \dots$ *there exists* T *such that* $\cap_{t=0}^{\infty} F_t' = \cap_{t=0}^{T} F_t$.

The theorem states that the outcome of the bargaining will always be a frame on the canonical path. The rivals would suggest more specific frames either until convergence or until a situation where any further specification would produce a frame that "misses their point," so-to-speak, by removing too much of the favorable outcome space for both players. Figure 3 shows a situation where parties could decide to stall on the current frame: If player 1 has to choose between retaining F_0, or playing F_1 which would result in the rival's playing F_2, player 1 might choose to remain in F_0 if the utility of any outcome of the subgame starting from F_2 is lower than that of F_0, as long as player 1 believes that player 2 would reason similarly.
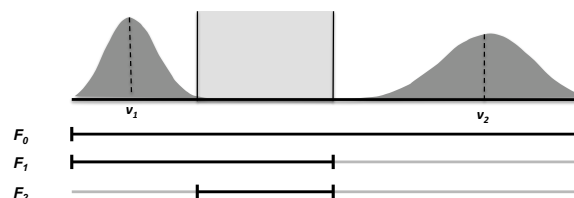

Figure 3: Stalled bargaining.

The idea of the proof is to construct a pair of strategies where each side attempts to pull the publicly accepted frame in the direction of its peak utility point. We show, assuming the peak of the first mover is to the left of peak of the second, that any deviation of the first mover would enable the second to shift the public frame more to the right, to an outcome of lower utility to the first mover. The full details of the proof of part 1 are given in the appendix; part 2 is proved in an accompanying technical report.

The equilibrium exhibits the following properties: (a) a first mover's advantage — for any player, the outcome would be closer to her peak point if she moves first than if she moves second; (b) a centrist's advantage — if a player moves first and her peak is closer to the middle of the initial frame, she can derive a higher utility from the outcome than if her peak were further from the middle. Please see appendix for justifications.

4 Discussion

4.1 Political communication

This article studies some properties of frame bargaining through metaphor in political communication, where rival politicians choose how to elaborate the current metaphor to educate the public about the ongoing situation in a way most consistent with their political preferences. Modeling the public preferences as highest relative entropy subset of possible states-of-affairs, we show that strategic choices by the politicians lead to a subgame perfect equilibrium where the less politically extreme player who moves first is at an advantage.

In a democracy, such player would typically be the government, as the bulk of voters do not by definition vote for extreme views, and since the government is the agent that brings about changes in the current states of affairs, and is thus the first and most prepared to explain them to the public. Indeed, Entman's model of frame activation in political discourse is hierarchical, with the govern-

ment (administration) being the topmost frame-activator, and opposition and media elites typically reacting to the administration's frame (Entman, 2003).

4.2 Metaphor in political communication

The role of metaphor in communication has long been a subject of interest, with views ranging from an ornament that beautifies the argument in the ancient rhetorical traditions, to the contemporary views of conceptual metaphor as permeating every aspect of life (Lakoff and Johnson, 1980).

In political communication specifically, metaphor has long been known as a framing device. Framing can be defined as "selecting and highlighting some facets of events or issues, and making connections among them in order to promote a particular interpretation, evaluation, or solution" (Entman, 2003). Metaphors are notorious for allowing subliminal framing, where the metaphor seems so natural that the aspects of the phenomenon in question that do *not* align with the metaphor are seamlessly concealed. For example, WAR AS A COMPETITIVE GAME metaphor emphasizes the glory of winning and the shame of defeat, but hides the death-and-suffering aspect of the war, which makes sports metaphors a strategic choice when wishing to arouse a pro-war sentiment in the audience (Lakoff, 1991). Such subliminal framing can often be effectively contested by merely exposing the frame.

Our examples show a different use of metaphor. Far from being subliminal or covert, the details of the metaphor, its implications, and the evaluation promoted by any given version are an important tool in the public discussion of a complex political issue. The function of metaphorical framing here resembles a pedagogical one, where rendering an abstract theory in physics (such as electricity) in concrete commonsensical terms (such as water flow) is an effective strategy to enhance the students' understanding of the former (Gentner and Gentner, 1983). The measure of success for a given version of the frame is its ability to sway the public in the evaluative direction envisioned by the author by providing sufficient *educational benefit*, so-to-speak, that is, convincingly rendering a good portion of a complex reality in accessible terms.

Once a frame is found that provides extensive education benefit, such as the EUROPEAN INTEGRATION AS TRAIN JOURNEY above, a politi-

cian's attempt to debunk a metaphor as inappropriate risk public antagonism, as this would be akin to taking the benefit of enhanced understanding away. Thus, rather than contesting the validity of the metaphoric frame, politicians strive to find a way to turn the metaphor around, i.e. accept the general framework, but focus on a previously unexplored aspect that would lead to a different evaluative tilt. Our results show that being the first to use an effective metaphor that manages to lock the public in its framework is a strategic advantage as the need to communicate with the same public would compel the rival to take up the metaphor of your choice. To our knowledge, this is the first explanation of the use of extended metaphor in political communication on a complex issue in terms of the agendas of the rival parties and the changing disposition of the public being addressed. It is an open question whether similar "locking in" of the public can be attained by non-metaphorical means, and whether the ensuing dynamics would be similar.

4.3 Social dynamics

This article contributes to the growing literature on modeling social linguistic behavior, like debates (Somasundaran and Wiebe, 2009), dating (Jurafsky et al., 2009; Ranganath et al., 2009), collaborative authoring and editing in wikis (Leuf and Cunningham, 2001) such as Wikipedia (Vuong et al., 2008; Kittur et al., 2007; Viégas et al., 2004). The latter literature in particular sees the social activity as an unfolding process, for example, detecting the onset and resolution of a controversy over the content of a Wikipedia article through tracking article talk[7] and deletion-and-reversion patterns. Somewhat similarly to the metaphor debate discussed in this article, Viégas et al. (2004) note first-mover advantage in Wikipedia authoring, that is, the first version gives the tone for the subsequent edits and has its parts survive for relatively many editing cycles. Finding out how the initial contribution constrains and guides subsequent edits of the content of a Wikipedia article and what kind of argumentative strategies are employed in persuading others to retain one's contribution is an interesting direction for future research.

A number of recent studies of the linguistic aspects of social processes are construed as if the

---

[7]a page separate from the main article that is devoted to the discussion of the edits

events are taking place all-at-once — there is no differentiation between early and later stages of a debate in Somasundaran and Wiebe (2009) or initial and subsequent speed-dates for the same subject in Jurafsky et al. (2009). Yet adopting a dynamic perspective stands to reason in such cases.

For example, Somasundaran and Wiebe (2009) built a system for recognizing stance in an online debate (such as pro-iPhone or pro-Blackberry on http://www.covinceme.net). They noticed that the task was complicated by concessions — acknowledgments of some virtues of the competitor before stating own preference. This is quite possibly an instance of debate dynamics whereby as the debate evolves certain common ground emerges between the sides and the focus of the debate changes from the initial stage of elucidating which features are better in which product to a stage where the "facts" are settled and acknowledged by both sides and the debate moves to evaluation of the relative importance of those features.

As another example, consider the construction of statistical models of various emotional and personality traits based on a corpus of speed dates such as Jurafsky et al. (2009). Take the trait of intelligence. In their experiment with speed-dates, Fisman et al. (2006) found that males tend to disprefer females they perceive as more intelligent or ambitious than themselves. Consequently, an intelligent female might choose to act less intelligent in later rounds of speed dating if she has not so far met a sufficiently intelligent male, assuming she prefers a less-intelligent male to no match at all.

Better sensitivity to the dynamics of social processes underlying the observed linguistic communication will we believe result in increased interest in game-theoretic models, as these are especially well suited to handle cases where the sides have certain goals and adapt their moves based on the current situations, the other side's move, and possibly other considerations, such as the need to address effectively a wider audience, beyond the specific interlocutors. A game theoretic explanation advances the understanding of the process being modeled, and hence of the applicability, and the potential adaptation, of statistical models developed on a certain dataset to situations that differ somewhat from the original data: For example, a corpus with more rounds of speed-dates per participant might suddenly make females seem smarter, or a debate with a longer history would

feature more, and perhaps more elaborate, concessions.

5 Empirical challenges

We suggested that models of dynamics such as the one presented in this article be built over data where entities of interest are clearly identified. This article is based on chapters 1 and 2 of the book by Musolff (2000) which itself is informed by a corpus-linguistic analysis of metaphor in media discourse in Britain and Germany. We now discuss the state of affairs in empirical approaches to detecting metaphors.

5.1 Metaphors in NLP

Metaphors received increasing attention from computational linguistics community in the last two decades. The tasks that have been addressed are explication of the reasoning behind the metaphor (Barnden et al., 2002; Narayanan, 1999; Hobbs, 1992); detection of conventional metaphors between two specific domains (Mason, 2004); classification of words, phrases or sentences as metaphoric or non-metaphoric (Krishnakumaran and Zhu, 2007; Birke and Sarkar, 2006; Gedigian et al., 2006; Fass, 1991).

We are not aware of research on automatic methods specifically geared to recognition of extended metaphors. Indeed, most computational work cited above concentrates on the detection of a local incongruity due to a violation of selectional restrictions when the verb or one of its arguments is used metaphorically (as in *Protesters derailed the conference*). Extended metaphors are expected to be difficult for such approaches, since many of the clauses are completely situated in the source domain and hence no local incongruities exist (see examples on the first page of this article).

5.2 Data collection

Supervised approaches to metaphor detection need to rely on annotated data. While metaphors are ubiquitous in language, an annotation project that seeks to narrow the scope of relevant metaphors down to metaphors from a particular source domain (such as train journeys) that describe a particular target domain (such as European integration) and are uttered by certain entities (such as senior UK politicians) face the problem of sparsity of the relevant data in the larger discourse: A random sample of the size amenable to human an-

notation is unlikely to capture in sufficient detail material pertaining to the one metaphor of interest.

To increase the likelihood of finding mentions of the source domain, a lexicon of words from the source domain can be used to select documents (Hardie et al., 2007; Gedigian et al., 2006). Another approach is metaphor "harvesting" – hypothesizing that metaphors of interest would occur in close proximity to lexical items representing the target domain of the metaphor, such as the 4 word window around the lemma *Europe* used in Reining and Lönneker-Rodman (2007).

5.3 Data annotation

A further challenge is producing reliable annotations. Pragglejaz (2007) propose a methodology for testing metaphoricity of a word in discourse and report κ=0.56-0.70 agreement for a group of six highly expert annotators. Beigman Klebanov et al. (2008) report κ=0.66 for detecting paragraphs containing metaphors from the source domains LOVE and VEHICLE with multiple non-expert annotators, though other source domains that often feature highly conventionalized metaphors (like *structure* or *foundation* from BUILDLING domain) or are more abstract and difficult to delimit (such as AUTHORITY) present a more challenging annotation task.

5.4 Measuring metaphors

A fully empirical basis for the kind of model presented in this paper would also involve defining a metric on metaphors that would allow measuring the frame chosen by the given version of the metaphor relatively to other such frames – that is, quantifying which part of the "integration is a train journey" metaphor is covered by those states of affairs that also fit Thatcher's critical rendition.

6 Conclusion

This article addressed a specific communicative setting (rival politicians trying to "sell" to the public their versions of the unfolding realities and necessary policies) and a specific linguistic tool (an extended metaphor), showing that the particular use made of metaphor in such setting can be rationalized based on the characteristics of the setting.

Various questions now arise. Given the central role played by the public gratification constraint in our model, would conversational situations without the need to persuade the public, such

as meetings of small groups of peers or phone conversations between friends, tend less to the use of extended metaphor? Conversely, does the use of extended metaphor in other settings testify to the existence of presumed onlookers who need to be "captured" in a particular version of reality — as in pedagogic or poetic context?

Considerations of the participants' agendas and their impact on the ensuing dynamics of the exchange would we believe lead to further interest in game theoretic models when addressing complex social dynamics in situations like collaborative authoring, debates, or dating, and will augment the existing mostly statistical approaches with a broader picture of the relevant communication.

A Proof of Existence of a Subgame Perfect Equilibrium

For a segment $[a,b]$ and $a \leq v_1 < v_2 \leq b$ let $U_1(x) = \phi(||x - v_1||)$ and $U_2(x) = \phi(||x - v_2||)$ be utility functions with peaks v_1 and v_2, respectively. For a history $h = \{F_0, \ldots, F_t\}$ where $F_t = [l_t, r_t]$, let $\sigma_1^*(h)$, player 1's move, be defined as choosing $F_{t+1} = [l_{t+1}, r_{t+1}]$ such that $|F_{t+1}| = \frac{|F_t|}{2}$, and r_{t+1} is as close as possible to v_1. σ_2^* sets l_{t+1} with respect to v_2 in a symmetric fashion. Since F_t shrinks by half every round, $\lim_{t \to \infty} l_t = \lim_{t \to \infty} r_t = x^*$, converging to a point. We now show (σ_1^*, σ_2^*) is an equilibrium by showing that neither player has a profitable deviation.

Notice that after the first round the subgame is identical to the initial game with F_1 replacing F_0, and the roles of players reversed. Player 2 had no influence on the choice of F_1, hence she has a profitable deviation iff she has a profitable deviation on the continuation subgame where she is the first mover. It thus suffices to show that the first mover (player 1) has no profitable deviations to establish that (σ_1^*, σ_2^*) is an equilibrium.

Since by definition σ_2^* always chooses an entropy maximizing segment, for player 1 to choose a non-entropy maximizing segment (more or less than half the length) amounts to yielding the round to player 2, which is equivalent in terms of the resulting accepted frame to a situation where player 1 chooses an entropy maximizing segment – the same one chosen by player 2. Thus we need to consider only deviations with entropy maximizing frames.

Step 1: Suppose σ_1' is a strategy of player 1 and let $F_0', F_1', F_2', \ldots$ be the sequence of frames on

the path corresponding to the pair (σ_1', σ_2^*). Let t_0 be the first move deviating from the equilibrium path, namely $F_{t_0} \neq F_{t_0}'$. We first show that F_{t_0-1} could not be (a) completely to the left of v_1 or (b) completely to the right of v_2. Suppose (a) holds. Then by definition $r_{t_0-2} = r_{t_0-1} < v_1$, and, inductively, $r_0 = r_{t_0-1} < v_1$; this contradicts $r_0 = 1$ that follows from $F_0 = [0,1]$. Possibility (b) is similarly refuted. Therefore, the only two cases for F_{t_0-1} with respect to v_1 are depicted in figure 4. Note that this implies $v_1 \leq x^* \leq v_2$.
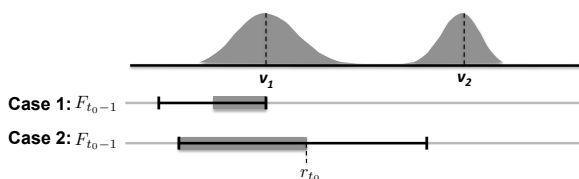

Case 1: F_{t_0-1}

Case 2: F_{t_0-1}

r_{t_0}

Figure 4: Two cases of current frame location.

Step 2: In case 1, σ_1^* will choose frames of type $[l_t, v_1]$ for any $t \geq t_0$, and σ_2^* will do the same on any history in the continuation game, hence the outcome will eventually be v_1. As this is player 1's peak utility point, she has no profitable deviation.

Step 3: In case 2, F_{t_0} is the leftmost entropy maximizing subsegment of F_{t_0-1} and the deviation F_{t_0}' can only be a shift to the right namely $r_{t_0}' \geq r_{t_0}$. If player 2 could choose $[v_2, r_{t_0+1}]$ given r_{t_0}, she can still choose the same frame given r_{t_0}', so the outcome would be v_2 and F_{t_0}' was not profitable. If player 2 could not choose $[v_2, r_{t_0+1}]$ given r_{t_0}, implying that $x^* < v_2$, but as a result of the deviation can now choose $[v_2, r_{t_0+1}']$, implying that the outcome would be v_2, clearly player 1 has not benefited from the deviation since U_1 is descending right of v_1. If player 2 still cannot choose $[v_2, r_{t_0+1}']$ after the deviation, she would choose the rightmost entropy maximizing segment with $l_{t_0+1}' \geq l_{t_0+1}$. If this still allows player 1 to do $[l_{t_0+2}', v_1]$ and hence to lead to v_1 as the outcome, it was possible in $[l_{t_0+2}, v_1]$ as well, so no profit is gained by having deviated. Otherwise, $r_{t_0+2}' \geq r_{t_0+2}$.

Step 3 can be repeated ad infinitum to show that $r_t' \geq r_t$ unless for some history h the deviation enables $\sigma_2(h) = [v_2, r_t']$. In the former case we get $\lim_{t \to \infty} r_t' = x' \geq x^* = \lim_{t \to \infty} r_t$ where $\cap_{t=1}^{\infty} F_t' = \{x'\}$. Since r_t' and r_t are to the right of v_1 and U_1 is descending right of v_1 it follows that $U_1(x^*) \geq U_1(x')$. In the latter case $x' \geq v_2$. Since F_t is never strictly to the right of v_2,

$x^* = \lim_{t \to \infty} l_t \leq v_2 \leq x'$, therefore $U_1(x^*) \geq U_1(x')$. In either case the deviation σ_1' cannot result in a better outcome for player 1. This finishes the proof that (σ_1^*, σ_2^*) is a Nash equilibrium.

Notice that (σ_1^*, σ_2^*) prescribe sub-strategies on any subgame that are themselves Nash equilibria for the subgames, hence (σ_1^*, σ_2^*) is a subgame perfect equilibrium $\square$

First Mover's Advantage: The proof of step 3 shows that having the left boundary of the current frame further to the right cannot yield a better outcome for player 1. Yet, if player 1's first turn comes after that of player 2, she will start with a current frame with the left boundary further to the right than the initial frame before player 2 moved, since moving the left boundary is player 2's equilibrium strategy. Hence a player would never achieve a better outcome starting second if both players are playing the canonical strategy.

Centrist's Advantage: Let M be the middle of F_0. Consider a more extreme version of player 1 — player $1^\#$. Suppose w.l.g. $v_1^\# < v_1 \leq M$. In case $v_1^\# < v_1 < v_2$, for all utilities u of the outcome of dynamics vs player 2, if player $1^\#$ could attain u, player 1 could attain u or more; the reverse is not true, for example when $|v_1^\# - l_t| < \frac{|F_t|}{2} \leq |v_1 - l_t|$ and player 1 (or $1^\#$) is moving first. In case $v_2 < v_1^\# < v_1$, if player 1 (or $1^\#$) moves first, she is able to force her peak point as the outcome. If $v_1^\# < v_2 < v_1$, player 1 can force v_1 as the outcome, whereas player $1^\#$ would not necessarily be able to force $v_1^\#$, as player 2 would pull the outcome towards v_2. Hence a first moving centrist is never worse off, and often better off, than a first moving extremist.

References

James Allan, editor. 2002. *Topic Detection and Tracking: Event-Based Information Organization*. Norwell, MA:Kluwer Academic Publishers.

Robert Aumann. 1990. Nash Equilibria are not Self-Enforcing. In Jean J. Gabszewicz, Jean-Francois Richard, and Laurence A. Wolsey, editors, *Economic Decision-Making: Games, Econometrics and Optimisation*, pages 201–206. Amsterdam: Elsevier.

Robert Axelrod and William D. Hamilton. 1981. The evolution of cooperation. *Science*, 211(4489):1390–1396.

John A. Barnden, Sheila R. Glasbey, Mark G. Lee, and Alan M. Wallington. 2002. Reasoning in metaphor

understanding: The ATT-Meta approach and system. In *Proceedings of COLING*, pages 121–128.

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2008. Analyzing disagreements. In Ron Artstein, Gemma Boleda, Frank Keller, and Sabine Schulte im Walde, editors, *Proceedings of COLING Workshop on Human Judgments in Computational Linguistics*, pages 2–7, Manchester, UK, August. International Committee on Computational Linguistics.

Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of non-literal language. In *Proceedings of EACL*, pages 329–336.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Resarch*, 3:993–1022.

Dwight Bolinger. 1980. *Language – The Loaded Weapon*. London: Longman.

Robin Clark and Prashant Parikh. 2007. Game Theory and Discourse Anaphora. *Journal of Logic, Language and Information*, 16:265–282.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 18:233–263.

Paul Dekker and Robert van Rooy. 2000. Bi-directional optimality theory: An application of game theory. *Journal of Semantics*, 17(3):217–242.

Anthony Downs. 1957. An economic theory of political action in a democracy. *The Journal of Political Economy*, 65(2):135–150.

James Druckman and Arthur Luria. 2000. Preference formation. *Annual Review of Political Science*, 2:1–24.

Robert M. Entman. 2003. Cascading activation: Contesting the White House's frame after 9/11. *Political Communication*, 20:415–432.

Dan Fass. 1991. Met*: a method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Raymond Fisman, Sheena Iyengar, Emir Kamenica, and Itamar Simonson. 2006. Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment. *Quarterly Journal of Economics*, 121(2):673–697.

Yoav Freund and Robert E. Schapire. 1996. Game theory, on-line prediction, and boosting. In *Proceedings of the annual conference on Computational Learning Theory*, pages 325–332, Desenzano del Garda, Italy, June -July.

Claire Gardent, Hlne Manuélian, Kristina Striegnitz, and Marilisa Amoia. 2004. Generating Definite Descriptions: Non-Incrementality, Inference and Data. In Thomas Pechmann and Christopher Habel, editors, *Multidisciplinary Approaches to Language Production*. Mouton de Gruyter.

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching metaphors. In *Proceedings of NAACL Workshop on Scalable Natural Language Understanding*, pages 41–48.

Deidre Gentner and Donald Gentner. 1983. Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner and A. Stevens, editors, *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Jacob Glazer and Ariel Rubinstein. 2001. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior*, 36(2):158–173.

Stephan Greene and Philip Resnik. 2009. More than Words: Syntactic Packaging and Implicit Sentiment. In *Proceedings of NAACL*, pages 503–511, Boulder, CO, June.

Avner Greif. 2006. *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. 2004. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501.

Andrew Hardie, Veronika Koller, Paul Rayson, and Elena Semino. 2007. Exploiting a semantic annotation tool for metaphor analysis. In *Proceedings of the Corpus Linguistics Conference*, Birmingham, UK, Julyt.

Jerry Hobbs. 1992. Metaphor and abduction. In Andrew Ortony, Jon Slack, and Oliviero Stock, editors, *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, pages 35–58. Springer Verlag.

Gerhard Jäger and Christian Ebert. 2008. Pragmatic Rationalizability. In *Proceedings of the 13th annual meeting of Gesellschaft fur Semantik, Sinn und Bedeutung*, pages 1–15, Stuttgart, Germany, September-October.

Dan Jurafsky, Rajesh Ranganath, and Dan McFarland. 2009. Extracting social meaning: Identifying interactional style in spoken conversation. In *Proceedings of NAACL*, pages 638–646, Boulder, CO, June.

Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. 2007. He says, she says: Conflict and coordination in Wikipedia. In *CHI-07: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 453–462, San Jose, CA, USA.

Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of NAACL Workshop on Computational Approaches to Figurative Language*, pages 13–20.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. Chicago University Press.

George Lakoff. 1991. Metaphor and war: The metaphor system used to justify war in the Gulf. *Peace Research*, 23:25–32.

Carlton E. Lemke and Joseph T. Howson. 1964. Equilibrium Points of Bimatrix Games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423.

Bo Leuf and Ward Cunningham. 2001. *The Wiki way: quick collaboration on the Web*. Boston: Addison-Wesley.

David Lewis. 1969. *Convention*. Cambridge, MA: Harvard University Press.

Robert D. Luce and Howard Raiffa. 1957. *Games and decisions*. New York: John Wiley and Sons.

Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic theory*. Oxford University Press.

Zachary J. Mason. 2004. CorMet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.

John Maynard Smith and George R. Price. 1973. The logic of animal conflict. *Nature*, 246(5427):15–18.

Nelson Morgan, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Adam Janin, Thilo Pfau, Elizabeth Shriberg, and Andreas Stolcke. 2001. The Meeting Project at ICSI. In *Proceedings of the HLT*, pages 246–252, San Diego, CA.

Andreas Musolff. 2000. *Mirror images of Europe: Metaphors in the public debate about Europe in Britain and Germany*. München: Iudicium.

Srini Narayanan. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of AAAI*, pages 121–128.

John F. Nash. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.

Douglass C. North. 1990. *Institutions, institutional change, and economic performance*. Cambridge University Press.

Ivandr Paraboni, Kees van Deemter, and Judith Masthoff. 2007. Generating Referring Expressions: Making Referents Easy to Identify. *Computational Linguisitics*, 33(2):229–254.

Prashant Parikh. 2001. *The Use of Language*. Stanford: CSLI Publications.

Keith T. Poole and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press.

Group Pragglejaz. 2007. MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1):1–39.

Kevin M. Quinn, Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2006. An automated method of topic-coding legislative speech over time with application to the 105th-108th U.S. Senate. Unpublished Manuscript.

Rajesh Ranganath, Dan Jurafsky, and Dan McFarland. 2009. It's not you, it's me: Detecting flirting and its misperception in speed-dates. In *Proceedings of EMNLP*, pages 334–342, Singapore, August.

Astrid Reining and Birte Lönneker-Rodman. 2007. Corpus-driven metaphor harvesting. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 5–12, Rochester, New York.

Ian Ross. 2007. Situations and Solution Concepts in Game-Theoretic Approaches to Pragmatics. In Ahti-Veikko Pietarinen, editor, *Game Theory and Linguistic Meaning*, pages 135–147. Oxford, UK: Elsevier Ltd.

Ariel Rubinstein. 1982. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.

Thomas C. Schelling. 1997. *The strategy of conflict*. Harvard University Press.

Reinhard Selten. 1965. Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit. *Zeitschrift für die Gesamte Staatswissenschaft*, 12:301–324.

Reinhard Selten. 1975. Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Form Games. *International Journal of Game Theory*, 4:25–55.

Shai Shalev-Shwartz and Yoram Singer. 2006. Convex Repeated Games and Fenchel Duality. In *Proceedings of NIPS*, pages 1265–1272.

Advaith Siddharthan and Ann Copestake. 2004. Generating referring expressions in open domains. In *Proceedings of the ACL*, pages 407–414, Barcelona, Spain, July.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing Stances in Online Debates. In *Proceedings of the ACL*, pages 226–234.

Robert van Rooij and Katrin Schulz. 2004. Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information*, 13(4):491–519.

Robert van Rooij. 2008. Games and Quantity implicatures. *Journal of Economic Methodology*, 15(3):261–274.

Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *CHI-04: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 575–582, Vienna, Austria.

John von Neumann and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton University Press.

Bernhard von Stengel. 2007. Equilibrium computation for two-player games in strategic and extensive form. In Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, pages 53–78. Cambridge University Press.

Ba-Quy Vuong, Ee-Peng Lim, Aixin Sun, Minh-Tam Le, and Hady Wirawan Lauw. 2008. On ranking controversies in Wikipedia: models and evaluation. In *Proceedings of the international conference on Web Search and Web Data Mining*, pages 171–182, Palo Alto, CA, USA.