Signaling games and non-literal meaning

Merlijn Sevenster

ILLC, UvA

June 7th, 2004

Outline of this talk

Main topics:

• Game theoretical notions help to model different linguistic phenomena

• Experimental results of Game theory shed light on the use of language

This talk:

- Introduction to signaling games
- Pay-off dominant equilibria
- *Super conventional signaling games*
- Risk dominant equilibria
- Experiments w.r.t. risk dominance
- Facts on SC signaling games
- Predictions
- Conclusion and future research

Signaling games

Quine (1936): How can meaning of language be conventionalized without presupposing meaning?

Lewis (1963): Consider meaning the result of playing *signaling games* rationally.

Though, Rubinstein: "[...] if game theory is to shed light on real life phenomena, linguistic phenomena are the most promising candidates. Game theoretical solution concepts are most suited to stable life situations which are "played" often by large populations of players."

Signaling games, extensively

Structure of the game:

First, Nature picks state $t \in T$

Second, sender S knowing t sends a message $m \in M$ to receiver R

Third, receiver R knowing only m performs an action $a \in A$

Payoff w.r.t. t, m, a*:*

Every state t calls for an appropriate action $f(t) \in A$:

$$\mathsf{u}_\mathsf{S}(t, m, a) = \mathsf{u}_\mathsf{R}(t, m, a) = \begin{cases} 1, & \text{if } a = f(t) \\ 0, & \text{if } a \neq f(t) \end{cases}$$

Nash equilibrium

A pair of strategies $\langle \mathbf{s}^*, \mathbf{r}^* \rangle$ is a *Nash equilibrium*, if for all strategies $\mathbf{s}$ and $\mathbf{r}$

$$\mathsf{U}_\mathsf{S}(\mathbf{s}^*, \mathbf{r}^*) \geq \mathsf{U}_\mathsf{S}(\mathbf{s}, \mathbf{r}^*)$$

and

$$\mathsf{U}_\mathsf{R}(\mathbf{s}^*, \mathbf{r}^*) \geq \mathsf{U}_\mathsf{R}(\mathbf{s}^*, \mathbf{r}).$$

Pay-off dominance

A pair of strategies $\langle \mathbf{s}^*, \mathbf{r}^* \rangle$ is a *pay-off dominant Nash equilibrium*, if for all Nash equilibria $\langle \mathbf{s}, \mathbf{r} \rangle$

$$\mathsf{U}_\mathsf{S}(\mathbf{s}^*, \mathbf{r}^*) \geq \mathsf{U}_\mathsf{S}(\mathbf{s}, \mathbf{r})$$

and

$$\mathsf{U}_\mathsf{R}(\mathbf{s}^*, \mathbf{r}^*) \geq \mathsf{U}_\mathsf{R}(\mathbf{s}, \mathbf{r}).$$

Lewis: The eventual pay-off dominant Nash equilibrium (*signaling system*) represents the conventional meaning.

Wärneryd (1993) gives a evolutionary characterization for pay-off dominant Nash equilibria.

Non-literal speech

Signaling system $\langle \mathbf{s}, \mathbf{r} \rangle$ accounts for meaning of $\mathbf{s}(T)$. But can only account for their *literal* meaning.

Substantial amount of speech is *non-literal*, e.g.

Metaphor:
"George Bush is a pig"

Irony:
"He is even more hansom than Brad Pitt"

Euphemism:
"Bill Gates is not very poor"

Typically a message m is used non-literally if it intends to communicate state t that is conventionally communicated by means of message m', where $m \neq m'$.

Non-literal speech is risky

Rewards of non-literal speech:

Social: politeness, face-saving, emphasizing and reinforcing claims to common ground

Cognitive: non-literal utterances are more deeply embedded in the audience's memory and have long-term effects that literal utterances have not

Efficiency

Risks of non-literal speech:

Social: Sally (2003): "A mismatch [...] between close [interlocutors] signals a problem with the relationship and may cause strong negative emotions and distancing"

Efficiency: parts of conversation have to be reconstructed

Risk dominance

In Game theory "risky equilibria" are modeled by notion of *risk dominance*, as opposed to pay-off dominance.

Harsanyi & Selten (1988): $\langle \mathbf{s}^*, \mathbf{r}^* \rangle$ is a risk dominant Nash equilibrium, if for all Nash equilibria $\langle \mathbf{s}, \mathbf{r} \rangle$

$$(\mathrm{U_S}(\mathbf{s}^*, \mathbf{r}^*) - \mathrm{U_S}(\mathbf{s}, \mathbf{r}^*))(\mathrm{U_R}(\mathbf{s}^*, \mathbf{r}^*) - \mathrm{U_R}(\mathbf{s}^*, \mathbf{r}))$$
$$\geq$$
$$(\mathrm{U_S}(\mathbf{s}, \mathbf{r}) - \mathrm{U_S}(\mathbf{s}^*, \mathbf{r}))(\mathrm{U_R}(\mathbf{s}, \mathbf{r}) - \mathrm{U_R}(\mathbf{s}, \mathbf{r}^*))$$

Typically, risk dominant equilibria provide better outcomes in worst-case scenarios.

	$\mathbf{r}^*$	$\mathbf{r}$
$\mathbf{s}^*$	2, 2	2, 0
$\mathbf{s}$	0, 2	3, 3

Two scenarios

Scenario A: Suppose you are playing the game with an arbitrary, unknown opponent.

Scenario B: Suppose you are playing the game with your best friend.

	$\mathbf{r}$	$\mathbf{r}'$
$\mathbf{s}$	10, 10	10, 0
$\mathbf{s}'$	0, 10	15, 15

What would you do?

Two rules of thumb

Harsanyi and Selten thought that players first coordinate on pay-off dominant equilibria. And that, if none are available, they coordinate on risk dominant equilibria. However, experimental Game theory has proven this conjecture false.

Rule 1: In a game with one outcome risk dominant and another "modestly" pay-off dominant, the former is more likely to be chosen.

Rule 2: As sympathy between the players increases, a pay-off dominant, risk dominated equilibrium is more likely to be realized.

Three facts

Fact 1 If $\langle \mathbf{s}, \mathbf{r} \rangle$ is a signaling system, then $\langle \mathbf{s}, \mathbf{r} \rangle$ is a Nash equilibrium

Fact 2 $\langle \mathbf{s}, \mathbf{r} \rangle$ is a pay-off dominant Nash equilibrium iff $\langle \mathbf{s}, \mathbf{r} \rangle$ is a signaling system and for every $t \in T$ it is the case that $\mathbf{s}(t) \neq \mathbf{cs}(t)$

Fact 3 If $\epsilon' > \epsilon$, then $\langle \mathbf{s}, \mathbf{r} \rangle$ is risk dominant iff $\mathbf{s} = \mathbf{cs}$ and $\mathbf{r} = \mathbf{cr}$.

Rule 1 and 2 applied

Rule 1: In a game with one outcome risk dominant and another "modestly" pay-off dominant, the former is more likely to be chosen.

Rule 2: As sympathy between the players increases, a pay-off dominant, risk dominated equilibrium is more likely to be realized.

Sally (2003): "[...] people play the language game in a way that is consistent with their play in all games."

Prediction *Rule 1*: Interlocutors communicate according to the convention, by default

Prediction *Rule 2*: As sympathy between interlocutors increases, the more likely they are to communicate non-literally.

Conclusion

- Solution concepts characterize linguistic phenomena

- Risk dominance is suited to model non-literal speech

- Game theoretical considerations concerning primacy of solution concept are of interest to pragmatics

Future research

• SC signaling games are not sensitive to metaphors, irony, euphemisms, etc.

• Formalization of the notion of sympathy/ common ground that seems crucial in *Rule 1* and *2*

• Risk dominance applied to other linguistic phenomena, such as the use of pronouns

• What solution concepts have what linguistic counterparts?