Agent Based Simulation Framework for Quantitative and Qualitative Social Research: Statistics and Natural Language Generation

Samer Hassan[1], Juan Pavón[1], Millán Arroyo[2], Carlos León[1]

[1] Dep. Ingeniería del Software e Inteligencia Artificial. Universidad Complutense de Madrid
samer@fdi.ucm.es, jpavon@sip.ucm.es, cleon@fdi.ucm.es
[2] Dep. Sociología IV, Universidad Complutense de Madrid.
millan@cps.ucm.es

Abstract. Even though Agent Based Social Simulation is beginning to be spread out as a powerful quantitative method for sociologists, it is still far from attracting qualitative ones. We propose to broaden ABSS horizons with a system that returns outputs useful for both paradigms. The case used as example is the study of the evolution of religiosity in the Spanish post-modern society. From a "macro" perspective, it analyses social trends, using quantitative data from the European Values Survey and giving evolution statistics. From the "micro" one, it can generate a narrative personal story of a representative agent, much closer to the Interpretativism tools. Several methods for choosing the most representative agent are commented.

Keywords: Social simulation, abss, mas, qualitative, quantitative, social research, natural language generation

1 Introduction

Sociologists and other social scientists use different research methods in order to study human societies. Social methods can generally be subdivided into two broad categories, even though both involve a systematic interaction between theories and data [4][9]. Quantitative methods, based on natural science and positivism, are concerned with attempts to quantify social phenomena and collect and analyse numerical data, with focus on the links among a small number of variables across many cases. For them, the social world is something that is *out there*, external to the social scientist, and waiting to be researched (objective vision). Qualitative methods, on the other hand, are based on comprehension, which emphasizes personal experiences and interpretation (subjectivism) over quantification, and is more concerned with understanding the meaning of social phenomena, with focus on links among a larger number of attributes across relatively few cases. Interpretivists believe that the social world is constructed by social agency and therefore any intervention by a researcher will affect social reality (interaction builds social reality).

However, it is increasingly recognized that the significance of these differences should not be exaggerated and that quantitative and qualitative approaches can be and must be complementary [6]. Quantitative methods are useful for describing social phenomena, especially on a larger scale. Qualitative methods allow social scientists to provide richer explanations (and descriptions) of social phenomena, frequently on a smaller scale. By using two or more approaches researchers may be able to 'triangulate' their findings and provide a more valid representation of the social world [10]. In fact, we can see more and more how sociologists of “one side” use tools from “the other side” in an assistant way.

There are already several works that show how social simulation can be a useful tool for quantitative researches [12]. From this point of view, agent based modelling even could be thought as an experimental tool for theoretical quantitative approach, a platform for empirical studies of social systems. With the aim of supporting this statement, we are working on a social simulation system for a specific sociological problem, trying to use it as a social research lab.

Qualitative researches have been used in this field just in an assistant way (in a second level), working from the quantitative perspective and letting it to guide the analysis and results view, as it can be seen in [16]. On the other hand, we want to propose, even knowing that the use of computers is mostly linked to quantitative researchers, the use of the results of agent based social simulation (ABSS) also for qualitative social scientists, at least in an assistant way. The approach is to consider ABSS not only for getting results that may be useful for a quantitative analysis of a social model, but for providing some assistance from a qualitative perspective. This is made by having advantage of the ability of agents to evolve autonomously and their interactions with the environment and other agents. By recording “stories” of agents, it is possible to get knowledge about their particular evolution and build a narrative personal story, which may provide insight on their motivation along time.

This work extends some experiments that have been performed by our research group with some sociologists of our University concerning the study of the evolution of religiosity in Spain [14]. In principle, the ABSS used for this case was applied to get quantitative data around a set of features that characterize the individuals. Here we are adding the generation of stories on some particular agents of the society, in order to get some qualitative argumentation. The next section presents roughly the multi-agent system that has been used for modelling the society. Then, section 3 presents a macro perspective of the analysis of the system, which would correspond to the quantitative approach. And next section goes to the individuals, the micro perspective, and explains how stories are generated and their use for qualitative research. At the end we provide some discussion on the potential of using ABSS for gaining knowledge on a society from the combination of qualitative and quantitative approaches.

2 Modelling of Social Systems with Multi-Agent Systems

Social phenomena are extremely complicated and unpredictable, since they involve complex interaction and mutual interdependence networks. Quantitative sociological

explanations deal with large complex models, involving many dynamic factors, not subject to laws, but to trends, which can affect individuals in a probabilistic way. A social system consists of a collection of individuals that interact among them, evolving autonomously and motivated by their own beliefs and personal goals, and the circumstances of their social environment.

The idea beneath ABSS is that we may be able to understand this huge complexity not by trying to model them at the global level but instead as emergent properties of local interaction among adaptive autonomous agents who influence one another in response to the influences they receive [12]. Because of that, the specification of characteristics and behaviour of each agent is critical, so it can manage the dimensions of the studied problem.

In the Multi-Agent System (MAS) designed, as explained in [14], agents have been developed with several attributes: from simple ones such as sex or age, to complex ones like for example ideology or educational level. The attributes have been chosen with respect to the context of the problem to solve, as we will see. The population in the agents' society (as in real societies) also experiments demographic changes: individuals are subject to some lifecycle patterns: they are born, get married, reproduce and die, going through several stages where they follow some intentional and behavioural patterns.

Moreover, agents/individuals can build and be part of relational groups with other agents: they can communicate with other close agents, leading to friendship relationships determined by a rate of similarity. Or, on the other hand, they can build family nuclei as children are born close to their parents. We can see a snapshot of the agents' space in Figure 1. The friendship relationships are represented by yellow links and families by green ones. Agents have different colour depending on their age and sex.

Taking the underlying sociological model, the parameters of the social simulation system fit all together logically. In this way, the system may be configured to reflect the parameters (such as average number of children per couple, or mean of male average age of death) from a specific country or even import data from surveys that specify the attributes of the agents, reflecting the behaviour of the given population.

Besides, due to the relative simplicity of the agents, the system can manage hundreds (and even thousands) of them, reaching the necessary amount for observing an emergent behaviour that results from the interactions of individuals, leading to the appearance of social patterns than can be studied [3]. And for this study, during and after the execution of the simulation tool several graphs may be plotted that reflect the evolution of the main attributes of the social system.

The system robustness has been tested enough to demonstrate the stability of the results, needed for the macroscopic analysis. Besides, the system has a deep diversity, attending to the differences between individuals. This fact will be useful for the micro analysis.

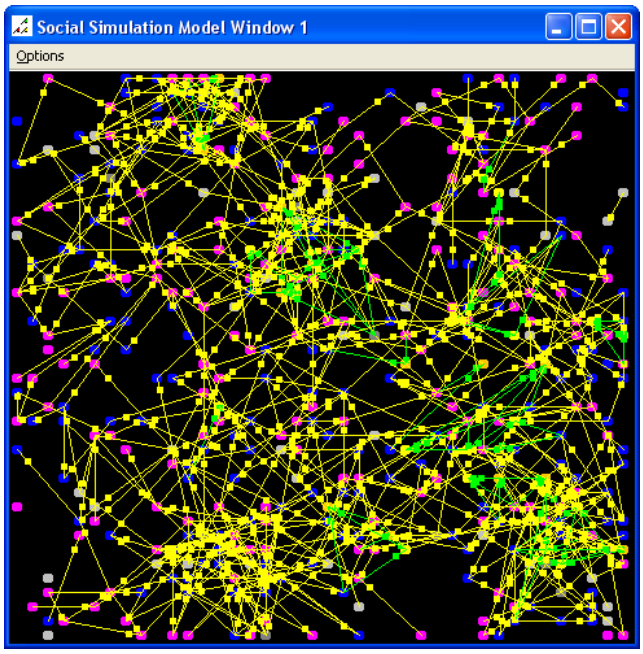

Fig. 1. Snapshot of the social simulation, after a certain period of time. We can see complex networks of friends and family nuclei all over the space.

3 The "Macro" Approach: Experimentation for Quantitative Research

The case under study, which is used to illustrate the approach in this paper, makes an analysis of the evolution of religiosity in Spain between 1980 and 2000. Initially, the aim was to assess the usability of ABSS tools for sociologists, who are not skilled in computer programming. Therefore, we looked for a real sociological issue, as the one of religiosity in society. The problem was how to model this social problem with agent concepts.

Initial data for the model and the simulation has been taken from results of the European Values Survey (EVS), which periodically make surveys in all European countries. EVS provides a source of quantitative information and periodical results offer data for validation of the simulation model [1]. The problem chosen and its simulation is widely explained in [14], so here we only focus on the main points. We just want to notice that religious people were divided into four types [2], as we can see in Table 1: ecclesiastical are clearly the most conservative, followed by low-intensity. Then, alternatives have a more modernized mentality, while the most modern and left-winged are those non-religious.

Taking the EVS-1980 of Spain, the sociologist was able to build an Excel spreadsheet for the characterization of 500 individuals, which statistically represent the Spanish population. These data were taken as input to generate a population of agents in the model, which was simulated for evolution for a period of 20 years, till year 2000. The similarity of results of the simulation with those of EVS-2000, as shown in Table 1 and Figure 2, allow to validate MAS model for this case, on the attributes under consideration (more specifically, those defining religious values). Note that as the system is non-deterministic, the graphical results have some variations at each execution. Then, Figure 2 should not be taken as static output. Anyway, as it was previously commented, the trends are always very similar, even though the exact data have some small comparison error. Therefore, the system executions have structural similarity, as defined in [7].

By using this kind of simulations it is possible to experiment on assumptions about the influence of attributes, relationships and interactions of agents in the evolution of societal values, from a quantitative point of view.

Table 1. Evolution of the religious patterns of the Spanish society, according to the European Values Survey.

	1980	*1990*	*1999*
Ecclesiastical	33	25	22
Low-intensity	22	26	23
Alternatives	14	17	19
Non-religious	31	32	35

4 The "Micro" Approach: Tool for Qualitative Research

Given the results that can be obtained by analysis of the emergent behaviour of the MAS under simulation, some explanations and theories could be also envisaged at the "micro" level, on the individuals. If we make a U-turn of perspective, we will try to overcome the limitations of the pure statistical analysis of characteristics from individuals (variables). To achieve this aim we will simulate their biographic behaviour (that follows some rules), taking each individual as a whole, with the holistic perspective of qualitative researchers. From this point of view, the evolution of each individual has much more importance, and instead of considering the global emergent trend, we focus on people lives. Each person is unique and unrepeatable, and therefore extremely important. All this statements are consistent with qualitative methods.

We have pursued this discursive line till the point of giving name and surname for each agent: now each one represents a person instead of a number or an ID (it's quite different "i214 died" than "Pablo Martínez died"). Also, we have given the agents the possibility of "living" events across their lives, events that could change their future decisions.

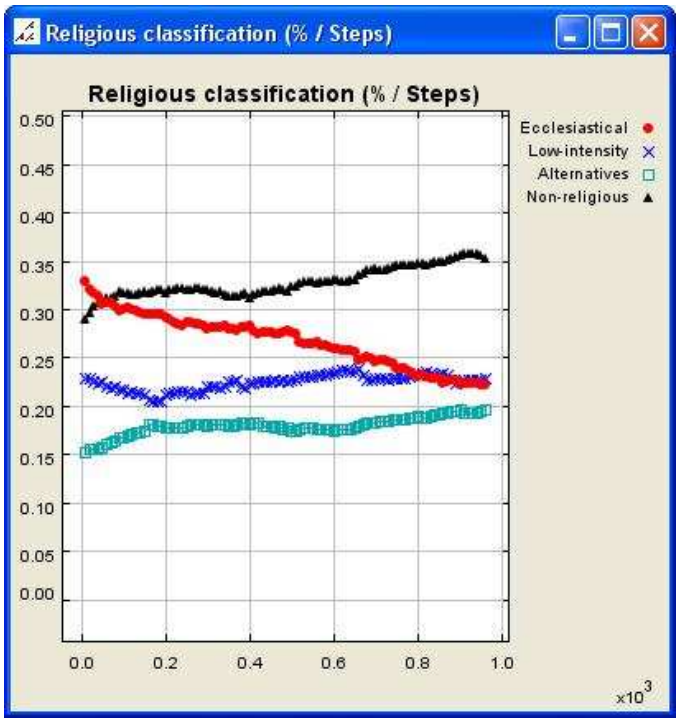

Fig. 2. Evolution of the religious patterns according to the social simulation system (1000 steps are equivalent to 20 years)

All these changes converge in the new output of the system: a narrated life-story in natural language. This is achieved by adding a new module to the system: a natural language processing (NLP) tool. This tool is a simple automatic narrator designed with most usual techniques on this area [15]. The main purpose of this system is to add more content to the analysis of the multi-agent simulation, and in this way completing the graphical output.

This Natural Language Generation System has been designed following simple and usual approaches based on rules. The system is heavily oriented on Content Determination (filtering the facts, telling only those facts that are considered to be really interesting for the reader), and Discourse Planning (ordering that filtered facts in such a way that the reader perceives a coherent story). Also very simple Surface Realization based on templates for creating the final text is addressed, to show a human readable form of the final content.

The most natural way of choosing one (or a few) individual stories to tell is to find the qualitative "ideal types". We want the life-stories of the most representative individuals, the ones that show what has happened during the simulation. These individuals reflect the "macro" changes of the complex society. For example, a hyper-inflation process can be shown through the fall of living quality of a representative

person. This person story would capture how he/she is forced to buy cheaper products and stop having some luxuries.

To achieve Content Determination, the NLP tool takes every individual as a whole being, and analyses the set of agents depending on its configured rules. Those rules, clearly context-dependant, have to be defined in order to choose correctly the representative individual(s). They should measure the interest (weight) of different life-events of the agents, and their relationships with others. In this way, only the most representative individual will be selected (the closest agent to the qualitative ideal type) from a set of many possible agents (500), the tool selects only one of them, that who is considered to be the most interesting one.

This filter is applied by computing a numerical value over every agent based on his life-events. This value, called *interest*, is based on a lookup table that assigns an heuristic value to each fact. Each of these values represent how much importance we give to each fact. Then the agents receives the *interest* value corresponding to the sum of the *interests* of their facts.

Then, with the *interest*, we apply some set of rules like filtering those agent whose interest falls below a given threshold, or removing redundant facts or relationships between agents.

Discourse Planning is performed by the use of templates that gives logical and coherent order to the text that has been filtered in the previous stage of Content Determination. The templates have simple sequences like time-ordering of the facts and describing the most important relationships.

To apply the rules of the original system to the social simulation domain it has been necessary to adapt the rules to this context. To perform this task we built a vocabulary with the aid of an expert sociologist in the domain. In this way we have built the vocabulary of the possible events and a logical path tree which defines the possible events that each agent can follow on each stage based on its own story (for example, a very religious woman will have a low possibility of deciding to abort). More details about how this tool works can be found on [11].

Biographies of agents that have been selected to build a story are not as expressive as real textual narrations, but can be directly compared with them (because of its natural language format). Due to the huge amount of textual and narrative material that qualitative research deals with, a narrative story is the best way to help them. This could help to understand the real individuals, compared with other possible life-stories, and it could be very useful due to the lack of individuals that usually qualitative researches have. If we use clustering or statistics for building the ideal types of the simulation, the real individuals of the qualitative research can be better understood compared with the simulated stories of these ideal types. Even we could use it to compare different qualitative researches over the same field.

Other applications of this new output could be to give some information very simple to understand for any human, not necessarily familiarized with the social simulation environment: for example an expertise on social behaviour. This expertise could read the text, that resumes with an example a life of some character, and extract information about the collective of people that have been simulated.

Moreover, this narration can be a good explanatory complement of quantitative researches, as other qualitative materials can be presented for supporting the final conclusions. This point will be deeply explained in next section.

For the example under study, as it is in a prototype phase, we have not followed a formal path for building the NLP rules that decided the most representative agent. We have decided the conditioned events and the implemented rules just using advises from an expert of the field. For computing this degree of representativeness, some simple formulas based on lookup tables has been applied, as can be seen in [11]. Even so, the generated text, in Figure 4, can give an idea.

The execution of the simulation of the MAS model, which resulted in quantitative macro results, generates also an XML file. This XML is the result of logging every event of every agent along their simulated lives. Agents that die and are born, matching or friendship relationships, life events: all is recorded. Next, the NLP module will process this XML, as explained in [11] and futher, to give the narrative output. We can compare Figure 3 and Figure 4 to check the text generation, although the big effort is to choose “Rosa Pérez” between the hundreds of individuals.

```
<Log Id="i49">
  <Description />
  <Attribute Id="name" Value="rosa" />
  <Attribute Id="last_name" Value="pérez" />
  <Attribute Id="sex" Value="female" />
  <Attribute Id="ideology" Value="left" />
  <Attribute Id="education" Value="high" />
        ...
  <Events>
  <Event Id="e1" Time="1955" Action="birth" Param="" />
  <Event Id="e2" Time="1960" Action="friend" Param="i344" />
  <Event Id="e3" Time="1960" Action="friend" Param="i439" />
  <Event Id="e4" Time="1961" Action="friend" Param="i151" />
  <Event Id="e5" Time="1962" Action="horrible" Param="childhood" />
  <Event Id="e6" Time="1963" Action="best friend" Param="i151" />
  <Event Id="e7" Time="1964" Action="believe" Param="god" />
  <Event Id="e8" Time="1964" Action="every week go" Param="church" />
        ...
  <Event Id="e16" Time="1968" Action="problems" Param="drugs" />
  <Event Id="e17" Time="1971" Action="grow" Param="adult" />
  <Event Id="e18" Time="1971" Action="friend" Param="i98" />
  <Event Id="e19" Time="1972" Action="involved" Param="labour union" />
  <Event Id="e20" Time="1972" Action="friend" Param="i156" />
  <Event Id="e21" Time="1973" Action="get" Param="arrested" />
  <Event Id="e22" Time="1973" Action="learn" Param="play guitar" />
  <Event Id="e23" Time="1975" Action="became" Param="hippy" />
  <Event Id="e24" Time="1976" Action="involved" Param="NGO" />
        ...
  <Event Id="e29" Time="1980" Action="impossible love" Param="i469" />
  <Event Id="e30" Time="1984" Action="couple" Param="i439" />
  <Event Id="e31" Time="1984" Action="live together" Param="i439" />
  <Event Id="e32" Time="1984" Action="child" Param="i679" />
  <Event Id="e33" Time="1984" Action="child" Param="i680" />
  <Event Id="e35" Time="1985" Action="friend" Param="i102" />
  <Event Id="e36" Time="1985" Action="divorce" Param="i439" />
  <Event Id="e37" Time="1987" Action="couple" Param="i102" />
  <Event Id="e38" Time="1987" Action="live together" Param="i102" />
  <Event Id="e39" Time="1987" Action="have" Param="abortion" />
        ...
  <Event Id="e48" Time="1995" Action="be unfaithful" Param="i102" />
  <Event Id="e49" Time="1995" Action="fired" Param="job" />
</Events>
</Log>
```

Fig. 3. XML example of life story

5 Conclusions and Further Work

We have seen the potential of the ABSS system for the quantitative analysis, reaching to the possibilities of social experimentation. We have experimented a way for expanding the possibilities of typical agent-based simulation, so they can build an output appropriate for the qualitative perspective.

```
Rosa Pérez was born in 1955, and she met Luis Martínez, and
she met Miguel López, and she met María Valdés, and she
suffered a horrible childhood, and she had a very good
friend: María Valdés, and she believe in God, and she went to
church every week, and she met David García, and she wanted
to be a priest, and she suffered an incredible accident, and
she met Marta Alonso.

When she was a teenager, she messed with a gang, and she met
Claudia Sánchez, and she went to confession every week, and
she had problems with drugs, and she became an adult, and she
met Marci Boyle, and she was involved in a labour union, and
she met Carla González, and she got arrested, and she learned
how to play the guitar, and she became a hippy, and she was
involved in a NGO.

She met Sara Hernández, and she stopped going to church, and
she met Marcos Torres, and she fell in love, desperately,
with Marcos Torres, but in the end she went out with Miguel
López, and she lived together with no wedding with Miguel
López, and she had a child: Melvin López.

She had a child: Andrea López, and she met Sergio Ruiz, and
she separated from Miguel López, and she went out with Sergio
Ruiz, and she lived together with no wedding with Sergio
Ruiz.

She had a abortion, and she bought a house, and she had a
depression, and she had a crisis of values, and she was
involved in a NGO, and she had a huge debt, and she inherited
a great fortune, and she met Daniel Lorenzo, and she bought a
car, and she was unfaithful to Sergio Ruiz with another man,
and she was fired from her job.

 Nowadays she is an atheist.
```

Fig. 4. The life story of a representative individual

With respect to the final appearance of the resulting text describing the life of one agent, it does not have yet good literary quality. It is necessary to generate more complex templates of text generation, as well as more sophisticated realization methods (those related with the final textual form of the narration). By defining better rules for choosing relevant agents, and for choosing most relevant facts of its life, the

final text will be much clearer for a human reading it. To achieve this aim we must test the different ways of building those rules (as were defined previously) and compare their results. These rules could be:

- Rules built from the ideal types of a previous qualitative research, guiding the analysis
- Rules built from statistics or by an expert in the field
- Rules built from clustering of the agent lives.

Furthermore, it is possible to create narrations about more than one character, by inserting parts of other agents' life. There is, of course, much work in progress about this, and, although natural language generation has many problems, it is possible to create much better descriptions of social agents.

To evaluate the results of the NLP system there are many possible approaches. In [8] an evaluation of the texts is presented. It could be interesting to compare this system with other representation systems based on qualitative approaches.

It is possible to use a conceptual ontology for the communication between the expert sociologist for representing the domain semantics. Thus, it would be possible to formalize the hierarchy of the facts and their meanings.

Also, we are considering enriching the kind of information that is extracted from agents' behaviour, about their mental state. This is specially interesting with agents that follow BDI model [5]. In this way, the qualitative aspects of the agents will be empowered, as well as their stories, completed with mental states and motivations (instead of only describing facts). If agents' interactions base on this model, the quantitative macro results obtained in this way would be much more directly connected to the micro behaviour. With this aimed achieved, we could a) use qualitative researches as input of the system, in the form of BDI data (attitudes, perceptions, behavioural patterns), as is done in [16]; or b) contrast our simulated biographies with qualitative empirical results, and validate the system model: if there is convergence between data, the guarantees of accuracy of the system will be bigger [13].

The application of the obtained stories as a complement for quantitative results can be quite useful. It would explain the "why" of certain phenomena in the social system: why they evolve in a way and not another. These questions usually arise in the social sciences when only quantitative data is analyzed, and are referred to phenomena derived from the complexity of individuals' interaction and system dynamics.

However, due to is relatively usual to 'triangulate' the methodological perspectives (as it was explained previously), we have to analyze which are the benefits of simulated biographies above real ones. They are faster, easier and cheaper to obtain (if we already have a simulated model), but the important issue of trust should be tackled. We have to test the simulation with known empiric data (mainly quantitative, but if it is possible, qualitative too) before we can consider its biographies trustworthy. Of course, if there are discrepancies, we should trust real data. But when both converge in the main aspects and there are some holes in real biographies that are covered with simulated ones, we could trust simulated results. This limitation can be assumed for the practice of social research, because it's usual to use interpretations and explanatory models that are not verified.

Acknowledgments. This work has been developed with support of Dirección General de Universidades e Investigación de la Consejería de Educación de la Comunidad de Madrid (Spain) and Universidad Complutense de Madrid (Grupo de investigación consolidado 910494) and the project TIN2005-08501-C03-01, funded by the Spanish Council for Science and Technology.

References

1. Andrés Orizo, F., Elzo, J.: España entre el localismo y la globalidad. La Encuesta Europea de Valores en su tercera aplicación, 1981-1999. SM, Madrid (2000)
2. Arroyo Menéndez, M.: Cambio cultural y cambio religioso, tendencias y formas de religiosidad en la España de fin de siglo. Servicio de Publicaciones de la UCM. Madrid (2004)
3. Axelrod R.: Advancing the Art of Simulation in the Social Sciences. Conte, Rosario., Hegselman, Rainer., and Terna, Pietro. (eds.). Simulating Social Phenomena, Berlin: Springer. (1997) 21-40.
4. Babbie, E.: The Practice of Social Research, 11th edition, Wadsworth, Thomson Learning Inc. (2007)
5. Bratman, M.E.: Intentions, Plans and Practical Reason. Harvard University Press. (1987)
6. Corbetta, P.: Social research: theory, methods and techniques. Sage, London, (2003)
7. Gilbert, N., K. G. Troitzsch: Simulation for the Social Scientist. Open University Press, Buckingham, U.K. (1999)
8. Hassan, S., Gervás, P., León, C., Hervás, R.: "A Computer Model that Generates Biography-like Narratives", 4th International Joint Workshop on Computational Creativity. University of London (2007)
9. Ibáñez, J.: "Del algoritmo al sujeto. Perspectivas de la investigación social". Madrid. Siglo XXI. (1985)
10. Jick, T.D.: Mixing qualitative and quantitative methods: triangulation in action. Administrative Science Quarterly, 24 (1979) 602-611
11. León, C., Hassan, S., Gervas, P.: From the Event Log of a Social Simulation to Narrative Discourse: Content Planning in Story Generation. AISB'07, Artificial and Ambient Intelligence, Proceedings of the AISB Annual Convention, Patrick Olivier and Christian Kray, Newcastle, UK (2007) 402--409
12. Macy, M.W., Willer, R.: From Factors to Actors: Computational Sociology and Agent-Based Modeling, Annual Review of Sociology, Vol. 28 (2002) 143-166
13. Moss, S., Edmonds, B.: Sociology and Simulation: Statistical and Qualitative Cross-Validation. American Journal of Sociology, volume 110 (2005), 1095–1131
14. Pavón, J., Arroyo, M., Hassan, S., Sansores, C.: Simulación de sistemas sociales con agentes software. Actas del Campus Multidisciplinar en Percepción e Inteligencia, Vol. I (2006) 389-400
15. Reiter, E., Dale, R.: Building Natural Language Generation Systems, Cambridge University Press, (2000)
16. Taylor, R.I.: Agent-Based Modelling Incorporating Qualitative and Quantitative Methods: A Case Study Investigating the Impact of E-commerce upon the Value Chain. Doctoral Thesis, Manchester Metropolitan University, Manchester, UK. (2003)