

## **A complexity science perspective on language spread**

Michał B. Paradowski

Institute of Applied Linguistics, University of Warsaw, Poland

<https://doi.org/10.36505/ExLing-2010/03/0035/000155>

### **Abstract**

Linguistic creativity is a manifestation of innovativeness. Investigating the diffusion of linguistic innovation requires an interdisciplinary approach, pooling competences from sciences humanistic, social, and strict (in particular mathematics and computational methods). The initial results will be presented of an empirical project analysing the social spread of neologisms in a microblogging site.

Key words: linguistic creativity, neologisms, innovation diffusion, social networks

### **A complex system: more than a complicated system**

A complex system must be distinguished from a complicated one. While no unanimously agreed definition exists of what constitutes a complex system, most researchers agree that in order to be classified thus, a system has to share most of the following properties (Paradowski 2009b):

- a large number of elements interacting via simple local rules;
- the emergent global behaviour of the system is not a simple product of the sum of the behaviours of its components (“more is different”);
- dynamic and evolving over time, with emergent properties;
- topological diversity, heterogeneity (and directedness);
- strong self-organisation (without an orchestrator), resulting in
- resistance to damage, resilience to failure and high flexibility due to
- learning and natural adaptability to changing conditions;
- ostensibly negligible perturbations and rare events on the local level lead to significant changes in the behaviour of the whole system;
- bifurcation and phase transitions;
- sensitivity to initial conditions;
- stability and multistability;
- hysteresis (with the consequence that the system is nondeterministic);
- scale-free – self-similarity;
- heavy-tailed distribution scaling like a power law;
- entropy (many possible ways of evolution);
- typically impossible to solve/verify/predict analytically – hence the fundamental role played by numerical simulation.

Examples of complex systems can be found at every step. In biology, ontogeny and phylogeny, the spot character of a cheetah, structure of the leaf, ant trails, termite mounds, flashing fireflies, chirping crickets, and

predator-prey ecosystems are all instantiations of complex systems. So are communities and social behaviours, from cooperation to conflict and riots. Complex are epidemics, nervous and immunological systems, neural networks, the climate and earthquakes, financial markets and price fluctuations, logistics and traffic jams. In the sphere of technology, the Internet, telecommunications infrastructure and power grids, and the World Wide Web instantiate complex networks. Last, but not least, the field where it all started – physics – has long been grappling with analyses and descriptions of non-equilibrium thermodynamics, crystals, boiling liquid, or the nuclear fission reaction.

In linguistics, within an individual, perceptual dynamics and categorisation in speech or word and sentence processing; across society, variations and typology, the rise of new grammatical constructions, semantic bleaching, language evolution in general, and the spread and competition of both individual expressions, and entire languages, are all inherently complex systems. Over 80 papers have been published dealing with language simulations. However, the vast majority of the *in silico* experiments carried out are inadequate to the scenery of the 21<sup>st</sup> c. (Paradowski 2009a). The models:

- only allow for Euclidean relationships (whereas nowadays more and more of our linguistic input covers immense distances);
- are ‘static’ (while mobility is not exclusively a 20<sup>th</sup> or 21<sup>st</sup>-c. phenomenon, as evidenced by warriors, refugees, missionaries, or tradespeople);
- assume a limited, identical number of ‘neighbours’ for every agent;
- presuppose identical perception of a given individual’s prestige by each of its neighbours, as well as
- invariant intensity of interactions between different agents;
- most fail to take into account multilingual agents;
- have no memory effect, and
- zero noise (while noise may be a mechanism for pattern change).

To address these limitations, rather than take a modelling outlook, we can tap into one repository of language data nearly perfectly suited to large-scale dynamic linguistic analyses – the Internet. After all, this medium stores data which is virtually unregulated, essentially uncensored, spontaneous, being immediately registered, interconnected, and amenable to relatively easy search and analyses with the use of statistical and concordancing tools.

### **Language on the Internet**

Hitherto research on language evolution and change focused on large time-scales, typically spanning at least several decades. Nowadays, observable changes are taking place much faster. According to the Global Language

Monitor a new English word is born roughly every 98 minutes. As a handful of popular recent expressions recall *alcopop*, *Ardi*, *bangster*, *brown state*, *choice mom*, *death panel*, *defriend*, *deleb*, *e-vampire*, *freemium*, *fundoo*, *funemployed*, *green washing*, *intexticated*, *jai ho*, *mobama*, *n00b*, *octomom*, *quendy-trendy*, *recessionista*, *seatmate of size*, *sexting*, *slumdog*, *teabagger*, *tramp stamp*, *wonderstar*, or *zombie bank*... Particularly useful for multi-angle analyses of language phenomena are Web 2.0 services, with content (co)generated by the users, especially the ones which allow enriching analyses with information concerning the structure of the connections and interactions between the participating users. Our empirical research project has set out to investigate how mutual relations and communication between Internet users impact the social diffusion of neological tags (semantic shortcuts) in a Polish microblogging site (since diffusion of linguistic innovation is a function of social structure; Rogers 1962/2003; Watts 2007).

Table 1. The microblogging site in numbers (at time of data dump)

Users	20k, over half logging on daily
Users in the giant component	5.5k (density 0.003)
Relations	110k
Tags	40k
Tagged statuses	720k

Predictably, the frequency distribution of tags scales like a power law, both for short (24-hr) and long (1-yr) timescales. The most interesting is tracking those tags which are idiosyncratic to the system (i.e., not used in offline discourse), where the only way to pick them up is from a neighbour in the site. This allows plotting the dynamics of the spread of the items in a closed circuit, relative to its structure and size.

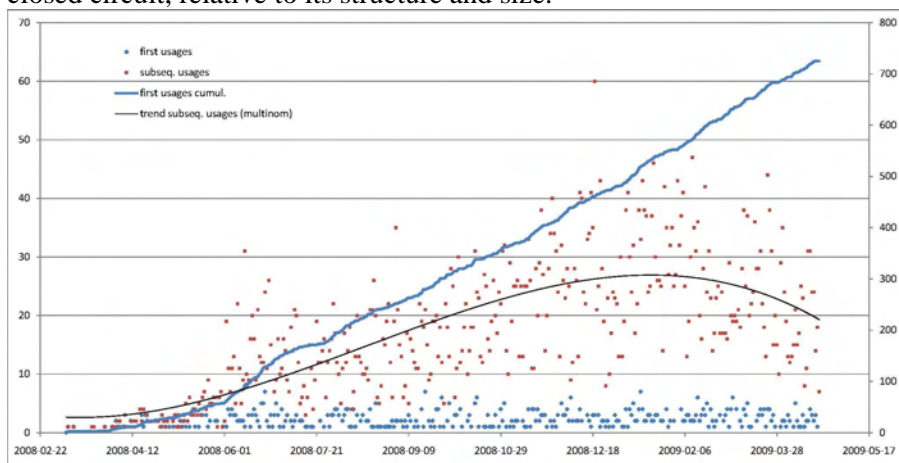


Figure 1. Usage of an idiosyncratic tag (Paradowski et al. 2010).

### **Social influence and diffusion**

The general innovativeness of Internet users scales not like a power law, but a hump-shaped unimodal (Paradowski, Jonak & Kuscsik 2009). The exposure thresholds specifying what proportion of one's neighbours must take up a certain activity for the individual to be likely to pursue it (in line with Granovetter's (1978) threshold model of collective behaviour and Valente's (1995) theory of network exposure) concentrate at low values, suggesting more susceptibility to social influence than may erstwhile have been expected, high innovativeness of the users, and 'virality' of tags. It may also mean that the innovative users are the ones with a greater number of subscriptions (Jonak & Paradowski 2009).

### **Acknowledgements**

The research has been supported by a grant from the Polish Society for Social Psychology, Agora SA, and Polskie Badania Internetu.

### **References**

- Granovetter, M. 1978. Threshold Models of Collective Behavior. *The American Journal of Sociology* 83(6), 1420–43.
- Jonak, Ł., Paradowski, M.B. 2009, Sep 16. Ekspansja słowa – analiza dynamiki rozpowszechniania się neologizmów leksykalnych i frazeologicznych oraz wzajemnych interakcji w społecznościach internautów. Invited presentation, VI scientific rally of the PSPS "Psychologia społeczna: teorie i badania", Lublin.
- Paradowski, M.B. 2009a, Jun 25. Complexity science and language phenomena – a successful marriage or a mésalliance? Invited talk, Complexity Science Seminar, Inst Math Sci, Imperial Coll London.
- Paradowski, M.B. 2009b, Oct 10. Applying a complexity science approach to analysing & modelling language phenomena. Invited lecture, Higher English Language Seminar, Dept English, Stockholm Univ.
- Paradowski, M.B., Jonak, Ł., Kuscsik, Z. 2009, Nov 23. The evolution of speech – towards network models of language development and spread. Poster, Intl Workshop "150 Years after Darwin: From Molecular Evolution to Language". Inst Cross-Disciplinary Phys & Complex Syst, Palma de Mallorca.
- Paradowski, M.B., Jonak, Ł., Kuscsik, Z. 2010, Jan 27–8. Follow the white rabbit, or how complexity science and linguistics can inform AI. 4<sup>th</sup> Intl Conference on Cognitive Systems, Inst Robotics and Intell Syst, ETH Zürich.
- Rogers, E.M. 2003. *Diffusion of Innovations*. New York, Free Press.
- Valente, T.W. 1995. *Network Models of the Diffusion of Innovations*. Cresskill, NJ, Hampton Press.
- Watts, D.J. 2007. The Accidental Influentials. *Harvard Business Review* 85, 22–3.