Waves in computer simulations of linguistic diffusion

by Luzius Thöny (University of Bern, Switzerland)

Abstract
Since the 19th century, the wave has been a key concept in descriptions of the diffusion of linguistic innovations in speaker communities. Recently, the study of language change through computer modeling and simulations has become more widespread. The aim of this paper is to investigate whether computer simulations of linguistic diffusion show wavelike phenomena. A brief review of diffusion mechanisms from previous literature is presented, followed by the description of a new software framework for testing hypotheses (PDIFFSIM) and a report of some results of own simulations. These results show that it is possible to implement a diffusion mechanism which produces waves of diminishing strength. Despite this, it is ultimately concluded that diffusion processes in the real world are better thought of as cascades in social networks rather than waves. Limiting factors of real world diffusion processes are rather to be sought in language-internal and -external restrictions.

(Running head: Waves in computer simulations)

Key words: language change, diffusion, agent-based modeling, wave, cascade, simulation, German

Introduction

In dialectology, descriptions of the diffusion of linguistic innovations in speaker communities frequently make reference to the concept of a wave (e.g. Wolfram & Schilling-Estes 2003: 713ff.). The wave is also at the core of ongoing discussions about models of language diversification (cf. e.g. François 2014). The *wave model* or *wave theory* goes back to the work of Hugo Schuchardt (1868: 34) and Johannes Schmidt (1872), who developed the idea at roughly the same time. The metaphor used is that of a stone thrown into a pond, which causes a wave to emerge at the point of impact. The wave then spreads outwards in a circular movement.

Schuchardt uses the metaphor of a wave in an attempt to capture the complicated relationships between language varieties–more specifically the Romance dialects he was working on–in a simple image. In an addendum to his work on the vocalism of Vulgar Latin (1868: 34), he writes:

> “Denken wir uns die Sprache in ihrer Einheit als ein Gewässer mit glattem Spiegel; in Bewegung gesetzt wird dasselbe dadurch, dass an verschiedenen Stellen desselben sich Wellencentra bilden, deren Systeme, je nach der Intensivität der treibenden Kraft von grösserem oder geringerem Umfange, sich durchkreuzen.” (Schuchardt 1868: 34)
>
> “Let us think of language in its unity as water with a smooth surface. It is set in motion through the emergence of centers of concentric waves at different locations, whose systems, which may be larger or smaller depending on the intensity of the force behind them, may cross each other” (my translation)

The wave model as a somewhat more elaborate theory was introduced shortly afterwards by Schmidt in his discussion of the genetic relationships of the Indo-European languages (1872). It was intended to replace the family tree model:

> “Wollen wir nun die verwantschaftsverhältnisse der indogermanischen sprachen in einem

> bilde darstellen, welches die entstehung irer verschiedenheiten veranschaulicht, so müssen wir die idee des stammbaumes gänzlich aufgeben. Ich möchte an seine stelle das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet." (Schmidt 1872: 27)
>
> "If we want to describe the relationships of the Indo-European languages using an image to illustrate the origin of their dissimilarities, we have to give up the notion of a family tree altogether. I would like to replace it with the image of a wave, which spreads outwards from the center in concentric circles and becomes weaker with increasing distance from the center." (my translation)

These descriptions allow us to establish three characteristics which define a wave according to Schmidt and Schuchardt:

1. the wave originates in a single point
2. it moves outwards from the point of origin in a circular movement
3. it becomes weaker with increasing distance from the center

Schuchardt's and Schmidt's descriptions of a wave of linguistic innovation are similar for the most part, but it is noteworthy to point out some slight differences: Schuchardt does not mention explicitly that the power of the waves diminishes as they travel outwards, while Schmidt does. Instead, Schuchardt emphasizes that waves can vary in their intensity, and this determines how far they reach. Regarding point (3) above, though, their position would appear to be essentially the same, because clearly even Schuchardt's waves need to diminish in power if the "intensity of force" is sometimes too low for two waves to cross each other.

The wave effects as described by the above-mentioned authors were difficult to investigate in the 19th century, and the situation remains the same today. Of course, we have much more data available nowadays, for example in the form of linguistic atlases, but the data tend to be synchronic snapshots and are therefore unsuitable for the study of diachronic language change. Unfortunately, a detailed empirical approach for studying wave phenomena as they happen in reality is out of the question. We would need to collect enormous amounts of data from very large groups of people over a long period of time. This is not a realistic option given the resources that are available for our research today. But there is an alternative to doing large scale empirical studies. We can use the processing power of modern computers to investigate virtual speaker communities, by designing stochastic models of speaker interactions and analyzing the results of computer simulations. This is the path that researchers in the field have followed. Since about the turn of the millennium, there has been an increase in studies aiming to simulate the diffusion of linguistic innovations in speaker communities using computers. Some examples are the studies by Nettle (1999a: Chapter 3, 1999b), Livingstone & Fyfe (1999), Baxter, Blythe & Croft (Baxter et al. 2006, 2009, Blythe & Croft 2012), the work done on self-organization in vowel systems by de Boer (2000, 2001, 2002), or the studies done by Steels (2011, Beuls & Steels 2013, among others), Stanford & Kenny (2013) and Pierrehumbert et al. (2014).

If we consider what has been said above about the wave model, such simulations raise an interesting question: the question whether these computer simulations show wavelike behavior. Investigating this point is the main focus of the present paper. In order to approach some answers, I will briefly review a number of diffusion mechanisms proposed in previous research. Following that, I will introduce a new simulation software and then report some results of my own simulations.

A brief review of diffusion mechanisms

An early attempt to describe the spread of an innovation as a wave that moves in time and space was made by R. A. Fisher in 1937. Fisher's model describes a genetic change in a population that spreads in geographical space under simple conditions. But the principle could be applied also to other situations: the spread of an epidemic, or a cultural phenomenon like a rumor or a technical innovation (cf. Cavalli & Sforza 1981: 40). The Fisher equation–a mathematical description of a wave moving at a certain speed from the origin–has been applied to archaeological and other problems, for example to the spread of agriculture throughout prehistoric Europe (Cavalli & Sforza 1981: 43). Note that this model of a "cultural wave" that spreads in geographic space has no specifically linguistic features, although it seems conceivable to apply it to linguistic topics as well. After all, Cavalli & Sforza's "wave of agriculture" mentioned above may already be an application to a linguistic topic, although indirectly, because the spread of agriculture is often associated with the advent of Indo-European languages in Europe[1]. However, its suitability to linguistic questions remains doubtful because of its strong simplifications. So far, I am not aware of any practical attempt to apply the Fisher equation to concrete problems in historical linguistics.

One of the earliest formalized linguistic wave models known to me is that by Bailey (1973: 67ff.). In his model, waves travel both through linguistic environments (68f.) and through social space (p. 70f.). The movement of the waves is highly simplified, as they appear to progress steadily and mechanically at each step in relative time.

More sophisticated diffusion models have been described in the literature since around the turn of the millennium. These studies generally operate in the framework of agent-based modeling (ABM; cf. Macal & North 2009 for an introduction and overview of the technique). The questions these authors tend to focus on are of the following kind: Under which circumstances does an innovation spread throughout the population and become fixed? Why do some innovations spread, and others do not? What does the trajectory of adoption look like? In fact, the issue of how an innovation can start to spread from a single point of origin and ultimately take over the entire speaker community has been a key question of previous simulation work.

This type of modeling work generally starts from the assumption that language change is usage-based, which means that speakers are affected by the frequencies of linguistic variants which they perceive. To illustrate the principle, let us take a single linguistic variable which could have either a value A or B. This could represent any feature of a language (or dialect) by which it differs from other languages (or dialects), e.g. a phonetic, lexical, syntactic or prosodic difference. Assuming that both variants A and B are present in the speaker community, it would seem intuitively plausible that an individual will over time adopt whichever one of the two variants he or she hears the most often (on the concept of interpersonal accommodation as a driving force of language change, cf. Auer & Hinskens 2005).

However, researchers who worked with computational models soon discovered that simple frequency-based accommodation does not lead to realistic patterns of diffusion. This has to do with the fact that an innovation–which by definition will always be outnumbered in the beginning–will never be able to get off the ground. The speaker who originally innovates will always be surrounded by neighbors who retain the conservative variant, and will immediately be converted back to it by these neighbors. This is a significant problem for the usage-based school of thought, and it is what Nettle refers to as the *threshold problem* (Nettle 1999b: 23.) The threshold problem can be overcome in computer simulations, but it requires that some kind of weighting or bias is used.

1 For some background on this (not undisputed) claim, cf. Mallory & Adams (2006: 452f.)

There are three basic ways to introduce a bias into the model. The first one is a social bias. It seems plausible that some speakers are more influential than others, for example because they have higher status in their society. They might also be more influential because of their position in the network, in that they have more connections and are therefore able to broadcast their personal language use to a higher number of other speakers.

A second type of bias could reside with the individual speaker and his or her willingness to adopt innovations. It seems reasonable to assume that some speakers are more innovation friendly and quicker to pick up new features than others. We can call this a speaker bias, because we assume that there is something special about an individual speaker that aids or hinders the adoption of an innovation.

The third possibility is a variant bias, sometimes called a functional bias (e.g. Nettle 1999b: 115). This type of bias is based on the idea that certain linguistic variants may have some inherent characteristics which make them more successful than others. An innovation might bring an advantage in any of the following areas:

1. ease of production,
2. ease of learning and recalling,
3. distinctiveness (higher distinctiveness means higher chance of successful communication),
4. conformance to systemic pressure (e.g. the trend towards symmetry in the vowel space),
5. social prestige and group identity.

We can compare such an advantage to the “fitness” of a gene in biological evolution. Similarly to how an advantageous gene provides some benefit to an individual, a linguistic variant may bring some advantage to a speaker that may lead to its wider spread.

All of the above-mentioned types of biases have been used in computer simulations, and applying them demonstrably leads to the successful diffusion of innovations under certain conditions (see below for examples). However, the results reported in previous studies suggest that such biases must be very strong to achieve the desired effect.

For example, Nettle (1999a: 48) operates with a social bias. He randomly assigns high status[2] to a minority of the agents and then, during language learning, gives the agents a bias towards the variants used by the high status agents. In Nettle's simulations, the introduction of such a social bias leads to an increase in linguistic diversity, i.e. to the survival of innovations. However, the social bias needed is very strong. A minority of 25% of the population is assigned high status in his scenario, and this minority appears to be almost exclusively responsible for determining which variants get selected. The case is similar in Nettle's experiments with the so-called Social Impact Theory (Nettle 1999b). Here, too, a strong social bias is used. The author states that sustained change occurs only in a scenario with so-called hyperinfluential individuals (p. 111, 114-6). The bias is introduced into the simulations by assigning each speaker a numerical value of their social status in the community, and then instructing the agents to pick up variants which are associated with high status. A small minority of agents–one in forty–is then assigned a status value which is much higher (by a factor of 25) than the highest value of the normal population (p. 111). That is an extreme bias by any measure, and what it would mean in reality is that a few people are immensely more influential than all the others when it comes to determining which variants succeed to spread

2 In Nettle's scenario, “high status” does not simply mean “upper class”, but is used as a more general construct for “whatever it is that makes someone an attractive role model” (p. 49 and footnote 3).

in a linguistic community.[3]

A different example of a model that uses a bias is described in a recent paper by Pierrehumbert et al. (2014). The crucial factor that makes the diffusion succeed in this model is the introduction of an innovation bias on the level of the individual. Each agent is assigned a bias either towards or against innovations. This means that some speakers are innovation friendly and pick up new features easily, while others are conservative and resist innovations. The scenario under which an innovation can spread throughout the network is then one in which an innovator happens to be connected to a number of highly innovation friendly speakers. In this situation, the seed of innovation will fall on fertile ground; it will be able to gain track quickly in a local pocket of the network and from there continue to spread even to more conservative speakers. Once again, a fairly strong bias is needed to achieve the desired result (p. 15-19). What this means in practice is that the scenario requires some radical early adopters who pick up innovations very easily, independent of what the innovation actually is. Furthermore, bias heterogeneity alone does not suffice to explain the successful spread of innovations, but the biases must also be distributed according to systematic patterns (p. 23).

On the matter of variant biases, it should be noted that at least in some cases of empirically observed language change, the successful innovation does not have any obvious advantage. For example, one dialect may show a diphthongization of long vowels, while another does not; or there may be subtle differences in word order of complex verbal constructions, or in intonation patterns. It is hard to see how one variant could be objectively advantageous compared to the other. Of course, we may simply not have identified the advantage yet, but considering the astounding variation found in languages and dialects around the world, it seems doubtful that a functional explanation could be found in all cases.

Another recent attempt at explaining the diffusion of linguistic innovations introduces the concept of momentum-based change (Stadler et al. 2014, 2016). The idea is that speakers have some attitude towards competing variants which affects their diffusion. More specifically, some variants may be perceived to be trendy or fashionable, while others may be perceived as antiquated. It could be precisely the fact that a feature is new and different from the current norm that makes it attractive. We may envisage the process of diffusion like this: At first, random fluctuations cause a variant to gain a certain level of momentum, at which point it is perceived as being fashionable. This momentum reinforces its wider spread and ultimately allows it to become the new norm. A challenge for this approach is the common claim in historical linguistics that speakers are often unaware of ongoing changes. For example, Nettle writes that “[t]hey [= people] are not even aware of most of the rules they effortlessly use or the linguistic changes in which they participate” (1999a: 13) and “[...] most linguistic changes and most linguistic variables are well below the level of conscious control” (1999a: 30). If that is the case, then it is hard to see how speakers could have a conscious attitude towards certain variants and how this could influence the course of linguistic change in any significant way.

I will conclude this brief overview of diffusion mechanisms by stating that current models do indeed demonstrate successful cases of the diffusion of linguistic innovations, but they require specific and strong biases. The next step is then to ask whether the diffusion processes in these models are indeed wavelike in the sense of Schmidt and Schuchardt.

3 Moreover, it seems somewhat inconsequential that Nettle allows for very large social biases and takes the results as an indication of the great importance of social factors, while at the same time stating that functional biases would appear to be “unlikely to be sufficient of themselves to allow rare variants to overcome the threshold problem and spread unless they are very large indeed.” (p. 114).

The weakening of waves

Let us return to Schmidt and Schuchardt's concept of a wave mentioned above. Three features were identified which characterize a wave: (1) it originates in a single point of origin, (2) it moves outwards in a circular movement, and (3) it becomes weaker as it travels away from the center. Of these, the first two are roughly consistent with current diffusion models. Innovations originating from one agent are possible, and the spread proceeds–in those cases where there is some spatial arrangement in (social) space–stepwise from the innovator, resulting in an (approximately) circular movement. But the third characteristic, namely the weakening of the wave as it moves away from the origin, does not appear to be a feature of any of the models I have reviewed so far. This point deserves to be treated in some more detail.

In Bailey's model (1973: 67ff.), waves appear to progress steadily and mechanically at each step in relative time. This leads to waves which gradually increase in strength. In fact, Bailey notes: "In contrast with the increasing attenuation of physical waves in time and space, the waves under discussion show increasing strength in time" (p. 79). According to my understanding, a change in Bailey's model, once it is initiated, will steadily progress until it reaches the edges of the network (i.e. it goes to completion).[4]

I see no indication in any of the more recent, agent-based simulation studies that the diffusion mechanism by itself would lead to a wave which diminishes in strength as it spreads. It therefore seems worth exploring whether a diffusion mechanism of this kind could be incorporated into the model. In the following sections, I will describe an attempt to implement such a diffusion mechanism, using a newly developed simulation framework called PDIFFSIM.

Own simulations using the PDIFFSIM simulation framework

The PDIFFSIM simulation framework is a simple software tool designed to test hypotheses about linguistic diffusion using agent-based modeling techniques. It was implemented using the Python programming language and makes optional use of the graph-tool library (Peixoto 2014). It can be run interactively with a visualization of the simulation shown on screen, or remotely on a server. In the latter case, visualizations can be stored as image files on the server's hard drive at any step of the simulation. Full source code of this framework is available online[5]. Simulations were run on a standard personal computer and on the Linux computer cluster of the University of Bern (UBELIX).

At the core of the simulation is the model of a speaker community which consists of a number of individual but interacting software agents. The agents are arranged on a two-dimensional grid. This is meant to represent a spatial arrangement in social space (this term is adapted from Bailey 1973: 13, 69). Distance in social space is not identical to geographical distance in reality, but it is related to it, as we are generally speaking more likely to communicate with people who are geographically close. A side-benefit of the grid arrangement is that it allows for easy visualizations of the simulations. For simplicity, agents only have a choice between two linguistic variants, one of which is the innovation and the other is the conservative variant. The innovation could be any characteristic by which two languages or dialects differ, e.g. a phonetic, lexical or syntactical difference. Each agent has a grammar g and a memory m. The grammar g is, in this heavily simplified model, a number between 0 and 1, i.e. the probability of uttering the innovation. The memory is a list of the variants which an agent has recently perceived. Giving the agents a probability of uttering the innovation is in line with the observation of dialectologists that speakers

4 However, Bailey allows for the reweighting of rules while the wave is propagated (Bailey 1973: 69), which might under certain circumstances hinder its progress.

5 The source code is hosted on GitHub: https://github.com/lvcivs/pdiffsim

do indeed produce variants of the same linguistic item (cf. Bailey 1973: 23ff., Seiler 2004: 380ff., Haas 2010: 662).

The simulation is initialized with a simple innovator ($g = 1$), whereas everyone else uses the conservative variant ($g = 0$). For each step of the simulation, the agents are paired up with a neighbor and they exchange a number of utterances. The utterances are produced according to the individual agent's grammar, and are appended to the memory of the other agent (the hearer). After the exchange of utterances, both agents adjust their grammar slightly depending on their recent memory.

The selection of a communication partner in the grid layout works according to a simple algorithm: The algorithm starts on the position of the current agent on the grid and moves one step up, right, down or left. If the agent at the new position is not a valid agent, a different one is chosen. If it is a valid agent, there is a 50% chance that this agent is selected as the communication partner and the algorithm stops, and a 50% chance to continue as above. After a maximum of 5 steps, the agent at the current position is selected as the communication partner. This algorithm has the effect that agents will always choose to talk to nearby agents, up to a maximum distance on the grid of 5 positions, but they are more likely to talk to close-by neighbors.

The central mechanism which allows the innovation to spread in this simulation is the adjustment that agents make to their grammar after each communication. By altering their grammar value slightly in order to approach what they perceive as the norm value, the agents accommodate to their environment by a kind of probability matching. The calculation for the accommodation process is given in formula (1):

$$g_i(t+1) = g_i(t) + \lambda \left(\frac{n_i}{m_i} - g_i(t) \right) \tag{1}$$

where t is the time step, g_i is the grammar of agent i, n_i is the number of occurrences of the innovation in the memory of agent i, m_i is the size of agent i's memory, and λ is the agent's adoption rate.

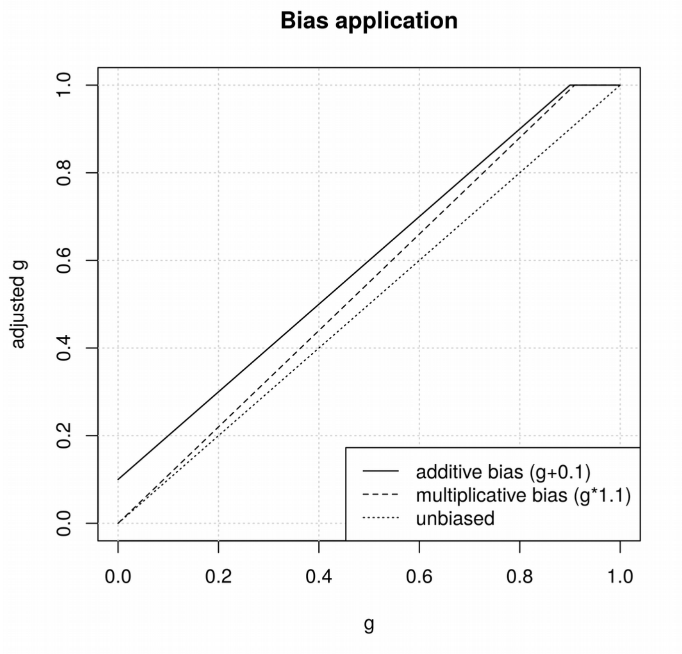
The accommodation formula in (1) leads to results comparable to those reported in previous simulation studies with so-called neutral evolution (Baxter et al. 2009: 270). In this scenario, it is possible but very unlikely that an innovation will spread from a single innovator. The following results from simulations with the PDIFFSIM framework may serve to illustrate this. A simulation was run 1000 times with a speaker community of 144 agents arranged in a grid layout and 500 time steps per simulation. The agents were fast adopting (λ=0.5) and no bias was used. These simulations yielded 53 runs (5.3%) where the innovation rose to a share of more than 5%, 2 cases (0.2%) where it rose to more than 20%, and zero cases where it rose above 30%.[6]

The bias introduced in my model is a variant bias.[7] The bias is applied at production time, which means that given a grammar g, the actual chance of this agent producing the innovation is an adjusted value of g (slightly higher than g). There are two ways in which the bias can be applied, and they have interesting consequences for the spread of innovations. The first way is to add a flat bias to the grammar, i.e. *p(innovation) = g + bias*. This has the effect that agents always have at

6 The share of the innovation is a measure of how widely the innovation has spread throughout the community. It was calculated by adding up the *g* values of all agents and dividing this by the total amount of agents, expressed as a percentage.

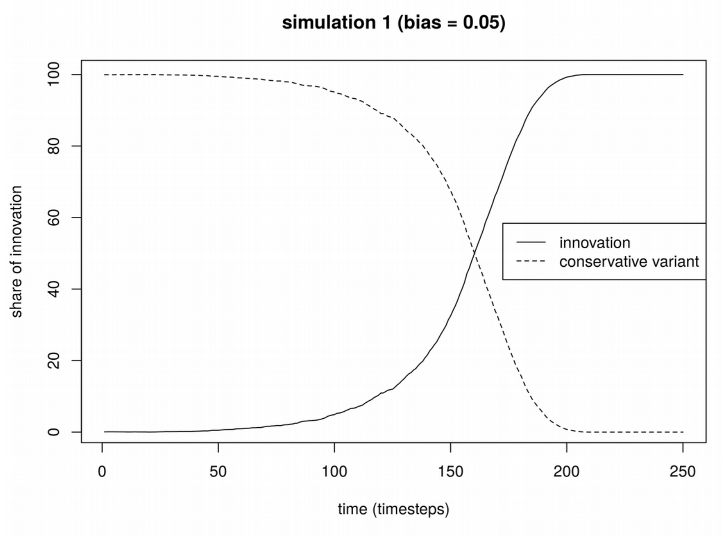
7 The emphasis on a variant bias is in line with the arguments of Seiler (2006) in favor of the importance of functional factors. Cf. also the evidence for functional factors in De Vogelaer (2006: 268, 270f.).

least a small chance to produce the innovation, even if they have never heard it. Alternatively, we can apply the bias by multiplication, i.e. *p(innovation) = g * (1+ bias)*. This alternative approach has two consequences: First, it means that the bias will have an effect only if the agent already has a chance to use the innovation, i.e. if $g > 0$. Second, it means that the effect of the bias increases as g increases. The difference between the two types of biases is illustrated in figure (1).



(figure 1)

For the simulations reported in this paper, a multiplicative bias has been used. This is because the additive bias leads to the innovation popping up randomly at various places on the grid, which results in decidedly un-wavelike patterns of diffusion. As we are looking for wavelike behavior in these simulations, it seems appropriate to use a multiplicative bias. Figure (2) illustrates how such a bias may lead to a successful diffusion of an innovation.[8]



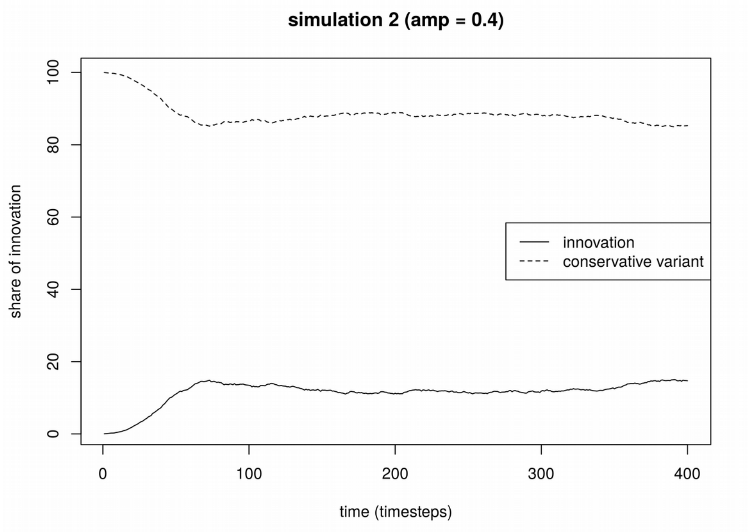
(figure 2)

8 This simulation was performed with a population of 1600 agents, $\lambda = 0.5$ and a variant bias of 0.05.

As seen in figure (2), the resulting *S*-curve is slightly skewed on the right hand side as an effect of the multiplicative bias. This means that the change is slow at the start, but goes to completion quickly after having gained traction.

Implementation of the wave mechanism

The new wavelike diffusion mechanism was implemented by introducing a variable into the model which represents the strength of the wave at a certain point. I call this the amplitude of the wave, as it can be thought of as the amplitude of the oscillation in a mechanical wave. The amplitude is coupled with the innovation and spreads along with it from agent to agent. It is strongest at the point of origin, and it is attenuated by a certain amount after each time step so that it will become weaker as it travels. The reduction in my implementation is by 5% after each time step, and the value is discarded when it falls below 0.01. When the agents produce utterances, they use the amplitude at their location as a bias in favor of the innovation. A high amplitude means a strong bias in favor of the innovation, while a low amplitude means a weak bias. Simulations run with this new diffusion mechanism do indeed show wavelike behavior. An example is shown in figure (3).[9]



(figure 3)

Figure (3) shows an initial phase where the innovation spreads along a trajectory which is roughly in the shape of an *S*-curve. But as the wave diminishes in strength, the growth gradually slows down and after some time, the wave effectively comes to a halt. From that point onwards the behavior is just that of a neutral evolution model (cf. Blythe & Croft 2012: 273f.).

Discussion

The waves as the one shown in figure (3) do in my opinion qualify as genuine waves in the sense of Schmidt and Schuchardt. They fulfill all three criteria, including the gradual weakening with increasing distance from the origin. But the crucial question is how realistic this scenario actually is. What gives rise to the waves in this simulation is that we pass along an additional value with the innovation while it spreads, namely the amplitude of the wave at that point. But what does this value correspond to in the real world? Do speakers keep track of how "strong" an innovation is, and do they pass on this value when they transmit the innovation to other speakers? It is not easy to believe that this is what happens in real speaker communities.

In fact, researchers who worked on social networks do not speak about a wave, but tend to talk

9 This simulation was performed with a population of 1600 agents, $\lambda = 0.5$ and a wave amplitude of 0.4.

about a cascade (e.g. Pierrehumbert et al. 2014: 5). The two concepts may seem similar at first, but there is an important difference. The cascade is a type of chain-reaction which does not naturally diminish in strength as it progresses. In the absence of any other limiting factor, it will keep going until the resources are depleted. Such a cascade is perhaps best illustrated by the domino-effect–the gradual collapsing of a row of standing domino stones after the first stone has been tipped over. As for a wave, how far it reaches depends on the impulse by which it was started. But for a cascade the initial impulse does not matter as long as it is strong enough to set the chain reaction in motion (for the domino stones, the energy that keeps the movement going comes from the transformation of potential energy, which the stones are storing due to their upright standing position, into kinetic energy as each consecutive stone tips over).

The metaphor of the wave clearly reaches its limits at this point. It may therefore be time to leave behind the idea of the wave, and start to think of a linguistic change in a speaker community as a cascade in a social network. Although it has been shown that it is indeed possible to implement a wave mechanism in such diffusion models, the design of that mechanism does not appear to represent a realistic scenario.

The abandonment of the idea of a wave, however, raises further questions, as there are empirical data of language change that suggest–at least at first sight–that waves of linguistic innovations do indeed become weaker as they travel. A case in point is the High German consonant shift which took place between the 5th and the 8/9th century AD in parts of the (pre-)German speaking area (Sonderegger 1979: 124ff.). The consonant shift is a complex set of changes in the consonant system of the language which involves a number of smaller sub-steps. The effects of the shift divide the German speaking area into multiple subareas, depending on how many of the sub-steps the local dialects participated in. The south, where the change is generally thought to have originated (p. 133f.), has the most changes. The areas further north have gradually less changes. Finally, the north has not been affected at all. What this pattern suggests is that we are dealing with a wave of innovation that came from the south, and became gradually weaker as it traveled northwards.

The only viable alternative interpretation is to adopt a view that we are not looking at a single diminishing wave, but instead at a set of waves which spread independently (consecutively) and which partially overlap each other. This view is also corroborated by the fact that the sub-steps can–to some degree–be traced as chronologically distinct processes in the written sources of the time (p. 128f.). Of course, nobody would doubt that the various sub-steps of the consonant shift are connected in some way, but that does not rule out the possibility that they *spread* independently. If we give up the idea of a wave, the latter view in fact seems more convincing. It seems preferable then to analyze the spread of the High German consonant shift not as a wave which diminishes in strength, but rather as a series of multiple, consecutive cascades which spread unequally far.

A different, but similar example is the diffusion of a set of vowel changes in Swiss German described by Haas (2010). He analyzes a lowering of short vowels as proceeding from the western areas of Switzerland, where we find all three sets of changes, through a transition area with two sets of changes, to the east with only one or no change. He also looks at the raising of low vowels, where the direction is the opposite, as we find most changes in the north-east, and gradually less towards the west. Once again, what appears at first sight like a wave of linguistic innovation which diminishes in strength as it spreads, may in fact be a series of cascades spreading independently of each other, and reaching unequally far.

An important point to note is that the "weakening" of a wave could actually be taken to mean two separate things. Firstly, it can mean that a wave becomes weaker in the sense that the innovation does not spread as quickly anymore, i.e. the amount of “converted” speakers per time unit decreases. This is actually the case in any of the described scenarios towards the end of the change, as the resources–in this case the “unconverted” speakers–inevitably run out as the innovation approaches completion. But this is not enough for real wavelike behavior, as the limiting factor is the amount of unconverted speakers which should be equal for any innovation. In other words, it does not explain why some waves travel further than others. Secondly, it can mean that the impact of the innovation diminishes, e.g. in the amount of linguistic contexts where it is applied. Both of the examples mentioned above, the second consonant shift and the vowel changes in Swiss German, are instances of the latter.

Although the idea of a cascade in a social network turns out to be a more fitting image for the spread of a linguistic change than a wave, the problem of the diminishing strength of the change remains unresolved. Cascades do not have any inherent feature that would make them slow down or stop before they have reached completion. This means that the empirical data from sound changes like the ones described above actually become harder to explain. This indicates that present diffusion models do not yet capture the essential parts of real diffusion processes. We may need to look beyond the usage-based diffusion mechanism of our (strongly simplified) models for restrictions that would have such an effect. A path to explore may be the insight that the adoption of an innovation can depend on the speakers' individual language system, in that the same innovation may be advantageous for some speakers but not for others. As an example, Haas (2010: 660) mentions the case of a sound change which will cause an (undesired) loss of a phonemic contrast in one variety, but not in another. It is easy to understand how the innovation could spread more easily in the second variety. Another example may be the filling of a morphological gap that exists only in a part of the linguistic area.[10] Linguists have also pointed to extra-linguistic reasons for a diffusion process to slow down or stop, e.g. when it reaches a language border or a cultural/political boundary (Haas 2010: 663f.).

Conclusions

The concept of a wave of linguistic innovation was introduced by Johannes Schmidt and Hugo Schuchardt in the 19th century. These authors used the wave as a metaphor to describe how linguistic innovations diffuse through speaker communities in time and space, and noted three characteristics of such a wave. A review of some of the models proposed in recent computational diffusion studies identified three different types of biases used to achieve successful diffusions: A social bias, an innovation bias, and a variant bias. Following that, it was noted that the third of the characteristics of a wave, the gradual weakening, is missing in these models, and a new mechanism was presented that is able to produce waves of that type. While these simulations demonstrated that it is possible to implement such a wave mechanism, it was ultimately concluded that this is not likely to be a realistic scenario. It seems more appropriate to think of the diffusion process as a cascade in a social network.

The question then arises how it is possible that a diffusion process sometimes slows down and stops rather than going to completion, as cascades do not have any inherent feature which makes them diminish in strength as they spread. But there are other factors which could have this effect, both of an intra- and extra-linguistic kind. Hopefully, more elaborate diffusion models will be able to take into account such factors in future simulation work. It may be possible to tackle some of these issues in the future by placing agents not in an abstract (social) space, but on a map with geographic features, as is commonly done in *Geographical Information Systems* (GIS; cf. Macal & North 2009: 93).

10 Cf. De Vogelaer 2006: 270 for a documentation of changes in the Dutch pronoun system where such an explanation may be appropriate.

Acknowledgements: This research was sponsored by the Swiss National Science Foundation, Grant. No. 151606. Most of the research was done during my stay at Stockholm University, Department of Linguistics, where I was generously hosted and advised by Prof. F. Lacerda during the period of the grant.

References

Auer, Peter & Frans Hinskens. 2005. The role of interpersonal accommodation in a theory of language change. In P. Auer et al. (eds.), *Dialect Change. Convergence and Divergence in European Languages*, 335-357. Cambridge: Cambridge University Press.

Bailey, Charles-James N. 1973. *Variation and Linguistic Theory*. Arlington: Center for Applied Linguistics.

Baxter, Gareth J., Richard A. Blythe, William Croft & Alan J. McKane. 2006. Utterance selection model of language change. *Physical Review* E 73:046118.1–20.

— 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21.257–96.

Beuls, Katrien & Luc Steels (2013). Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement. *PLoS ONE* 8(3).1–15.

Blythe, Richard A. & William Croft. 2012. S-Curves and the mechanisms of propagation in language change. *Language* 88 (2).269–304.

Cavalli-Sforza, Luigi L. & Marcus W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28.441–465.

— 2001. *The Origins of Vowel Systems*. Oxford: Oxford University Press.

— 2002. Evolving Sound Systems. In A. Cangelosi & D. Parisi (eds.), *Simulating the Evolution of Language*, 79–97. London: Springer.

De Vogelaer, Gunther. 2006. Actuation, Diffusion, and Universals. Change in the Pronominal System in Dutch Dialects. *Zeitschrift für Dialektologie und Linguistik* 73(3).259-274.

Fisher, Ronald A. 1937. The Wave of Advance of Advantageous Genes. *Annals of Eugenics* 7(4).355-369.

François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. In C. Bowern & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 161-189. London: Routledge.

Haas, Walter. 2010. A Study on areal diffusion. In P. Auer & J. E. Schmidt (eds.), *Language and Space*, Vol. I, 649-667. Berlin: de Gruyter.

Livingstone, Daniel & Colin Fyfe. 1999. Modelling the Evolution of Linguistic Diversity. In D. Florean (ed.), *Advances in Artificial Life: 5th European Conference: Proceedings*, 704–8. Berlin: Springer.

Macal, Charles M. & Michael J. North. 2009. Agent-based Modeling and Simulation. In M. D. Rossetti et al. (eds.), *Proceedings of the 2009 Winter Simulation Conference*, 86-98. Austin: Omnipress.

Mallory, James P. & Douglas Q. Adams. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: Oxford University Press.

Nettle, Daniel. 1999a. *Linguistic Diversity*. Oxford: Oxford University Press.

— 1999b. Using Social Impact Theory to simulate language change. *Lingua* 108.95–117.

Peixoto, Tiago P. 2014. The graph-tool python library. *figshare*. DOI: 10.6084/m9.figshare.1164194. URL: http://figshare.com/articles/graph_tool/1164194

Pierrehumbert, Janet B., Forrest Stonedahl & Robert Daland. 2014. A model of grassroots changes

in linguistic systems. *ArXiv* 1408.1985.1–30.
Schmidt, Johannes. 1872. *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
Schuchardt, Hugo. 1868. *Der Vokalismus des Vulgärlateins*. Vol. III. Leipzig: B. G. Teubner.
Seiler, Guido. 2004. On three types of dialect variation and their implications for linguistic theory. Evidence from verb clusters in Swiss German dialects. In B. Kortmann (ed.), *Dialectology meets Typology. Dialect Grammar from a Cross-Linguistic Perspective*, 367-399. Berlin: Mouton de Gruyter.
— 2006. The role of functional factors in language change. An evolutionary approach. In O. N. Thomsen (ed.), *Competing Models of Linguistic Change. Evolution and Beyond*, 163-182. Amsterdam: Benjamins.
Sonderegger, Stefan. 1979. *Grundzüge deutscher Sprachgeschichte. Diachronie des Sprachsystems*. Vol I. Berlin: de Gruyter.
Stadler, Kevin, Richard A. Blythe & Simon Kirby. 2014. Momentum-based language change: A non-adaptive model of directional selection. In E. A. Cartmill et al. (eds.), *The Evolution of Language. Proceedings of the 10th International Conference (EVOLANG10)*, 525-526. Hackensack, NJ: World Scientific.
Stadler, Kevin, Richard A. Blythe, Kenny Smith & Simon Kirby. 2016. Momentum in language change: A model of self-actuating S-shaped curves. *Language Dynamics and Change* 6(2).171-198.
Stanford, James N. & Laurence A. Kenny. 2013. Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change* 25(2).119–153.
Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8.339–56.
Wolfram, Walt & Natalie Schilling-Estes. 2003. Dialectology and Linguistic Diffusion. In B. D. Joseph & R. D. Janda (eds.), *Handbook of Historical Linguistics*, 713–35. Malden, MA: Blackwell.