1. Introduction

Research on Genres has developed a tradition of nearly half a century both in German and Anglo-American Linguistics. Research approaches are well-developed for offline and even TV genres. For Internet genres, the picture looks different. Crowston and Kwasnik, for instance, point out:

New document genres have emerged, while older ones have blended, changed, and been incorporated into different social endeavors [...] new electronic genres emerging frequently, appear to be shuffled, disassembled, and then put together again, in a seemingly chaotic manner. (Crowston & Kwásnik 2005: 79)

I think this quote is enough to show: The dynamic and heterogeneous realm of the WWW constitutes a special challenge. Whatever genre pattern you find today may be gone, blended or changed tomorrow. Genre research on the Internet means trying to grasp what does not want to be grasped.

However, users deal successfully and mainly intuitively with the new diversity every day. Therefore, it seems reasonable to assume a minimum degree of stability under the fluid surface. Stability is needed for the communication not to fail. That’s why my doctoral thesis is guided by the following questions:

1. Given the dynamics of the WWW – how can we grasp and describe whatever genre there may be on the Internet?
2. What are, diachronically, the forces of stability under the fluid surface?
I think: These questions call for a new angle for Internet genre research. Instead of genre fishing in today’s WWW, I propose a diachronic approach.

In my doctoral thesis, I am currently trying to develop such an approach with a special focus on personal weblogs. They can be roughly described as public online journals.

I chose Personal Weblogs for the following reasons:

- They constitute a genre with quite a long tradition of roughly 15 years on the WWW. We can assume both implicit and explicit norms to be established. Their genre history quite probably also involves offline genres such as the diary.
- There are already several studies on weblogs. However, no diachronic linguistic approach has been taken by now despite their long history.
- Last but not least, individual texts are relatively accessible diachronically compared to E-Mail- or Chat-genres, for instance.

In the next minutes, I will give an outline of my diachronic approach and its methodological considerations. Examples taken from my work in progress will illustrate the possible outcomes of the approach.

2. A Dynamic Genre Model

The particular dynamics of Internet genres calls for a flexible and dynamic genre model. I tried to develop such a model as a fusion of several branches of genre research. Let me sum up the most important points:

1. German text linguistics. In my model, I follow the concept of Textsorte that has been well established in German text linguistics. Genres (or Textsorten in the
German terminology) are situated on a fairly low level of abstraction as for instance the famous cooking recipe. Following Heinemann & Heinemann and others, Genres (or Textsorten) can and have to be described as a combination of several levels. Heinemann & Heinemann propose the levels of situation, function as well as semantic and linguistic structure. The cooking recipe, for instance, can be described in terms of its typical situation of use, its mainly instructive function and its typical choice of language.

2. I have extended this multi-layered approach using models taken from media linguistics. Media linguists propagate the concept of a Form of Communication such as Blog, E-Mail or Letter influencing embedded text types. On the Internet, the form of communication takes the shape of a software environment which sets the frame for text production. Texts on Twitter, for instance, cannot be longer than 140 characters due to the underlying software.

Media linguists also stress the importance of multimodal analyses. Genres are more than language, they include hypertextual structures, layout and other layers.

3. The Anglo-American Rhetorical School views genres as patterns of social action that are owned by a community. It stresses the importance of genre classifications present within the community. Those classifications can be explicit in that individual texts are declared as member of a genre; they can also be implicit when community members silently follow the established norm. Especially on the dynamic WWW, we need to rely on the virtual communities’ explicit declarations of genres to explicate their genre knowledge.

This social genre knowledge is prototypically organized, which includes fuzzy edges and partial overlap of genre categories. Individual texts can be seen as more
or less typical members of the category. The category can change when people start picking up innovations introduced by other net users, e.g. bloggers. So each text realizes the category and has a back effect on the category at the same time.

This model is adjusted to both the dynamics and the multi-modality of genres on the Internet. It allows for the fine grained analyses I want to carry out in my thesis.

3. An approach to diachronic genre research

3.1. The Internet Archive

A fundamental problem of diachronic genre research is the question of how to assemble a diachronic research corpus. To my knowledge, there are currently no diachronic Internet corpora available.

One way out of this problem is offered by the Internet Archive. Dating back to 1996, snapshots of the Internet have been collected. Via the Wayback Machine, the Internet Archive makes it possible to surf the Internet both synchronically and diachronically. That makes the Internet Archive a powerful database for diachronic research. However, there are some important points that have to be considered. The most important ones are summed up on your handout:

- The Wayback Machine of the Internet Archive currently only allows looking up definite URLs. That is, you have to know the address of the site you want to visit.
- So called dynamic documents such as search engines cannot be archived. The search engine google, for instance, turns useless when accessed via the Internet Archive.
• The software of the Internet Archive follows the Standard for Robot Exclusion. This document states that site owners can prevent their domains from being archived. This fact entails quite a high percentage of especially personal sites that cannot be found in the archive and generate error messages.

• Not every site has been archived equally well. Sometimes, a site can only be accessed partly for a particular date because of problems in the archiving process.

These limitations turn working with the archive into looking for the famous needle in the haystack. Nevertheless, the Internet Archive provides a powerful means for diachronic research. That’s why I used it as a source to assemble my research corpus.

3.2.  **DIABLOK – Diachrones Blog Korpus**

I termed my research corpus **DIABLOK**, which stands for **Diachrones Blog Korpus**. As I said, I worked my way through the Archive to assemble it.

The research corpus consists of two parts: For the earliest years from 1997-2000, there is only a handful of blogs. They were listed by Jessie James Garrets on his blog *Infoshift* under the heading *weblogs* in 1999.

For the second part, I spotted a blogging community called *Globe of Blogs*. It has been in existence since 2002. The community provides the opportunity of registering blogs in several categories. Those categories can be assumed to constitute the community’s inventory of blogging genres. One category is called *personal weblogs*. I accessed this category and collected sites for three periods. The periods were set according to the availability of texts in the *Internet Archive*. (Handout!)
Between part I and part II, there is a gap of one year. This gap is the result of the fact that *Globe of Blogs* was only founded in 2002 so that no *Personal Weblogs* could be collected for 2001.

Together, the two parts make up the *DIABLOK* – the first diachronic Internet corpus.

4. Applying the dynamic genre model– preliminary results and further research questions

In the last section, I would like to sketch roughly some preliminary findings that were generated by applying the genre model to the corpus.

4.1. General Change with respect to one feature

A first glance at the corpus already shows a general tendency of change: While early blogs have fairly short postings, the postings seem to get longer in more recent blogs. This can be seen in the examples: Quite short posts in the *Infoshift* blog from 1998; longer postings in *Ideal Life*.

A first test seems to support this observation: I randomly selected three blogs from each period of the corpus. I counted the words for the first up to 10 postings of each text. The figures for the average posting length seem to confirm the tendency to longer posts especially in the youngest blogs. However, each blog shows a high internal variance; the variety among the single blogs is fairly considerable, too. This trend needs to be further investigated in the more detailed analyses of my thesis.

The dynamic genre model helps to conceptualize this general change as a gradual movement of the category’s core. This movement mirrors a shift in the community’s
consensus on the category. The category’s variety expands as well as innovations take time to spread and may not reach all of the community’s members.

While the change of one isolated genre in one or more features is one possibility, it is quite likely that more often than not other genres are involved as well.

4.2. Genre Split

As in the case of genre splits: Starting from a relatively stable category, another category develops out of the first one. This second category exhibits other features on at least one dimension. Users start recognizing the second category as distinct from the first one.

Scott Rosenberg, one of the key bloggers, describes how blogs developed out of personal homepages when authors tried to find a suitable way of separating old and new material. Several steps were taken until the reverse chronological order – newest entries first – became established. This order is typical of blogs today. It is one important feature separating blogs from personal homepages.

According to the dynamic model, it is reasonable to assume a stage of transition during which recipients and authors are unsure of how to categorize certain texts. One example is provided by Joe Clark’s blog from 1999. The Infoshift Blog lists the site as weblog. The site itself however, is entitled “homepage” and contains a section called “weblog links of the week”. The table of contents also includes a section “five newest additions”, which is a typical way of personal homepages to point out new material. In sum, Joe Clark’s site is called both weblog and homepage and exhibits features of both. I would like to argue that it can be viewed as a transitory text.
4.3. Pattern Embedding

Other patterns of change can be subsumed under *Pattern Embedding*. In Pattern Embedding, bits and pieces of other genres are integrated into one target genre. The recipients perceive the result as member of one genre containing elements of another. Pattern Embedding can be described on several layers. The most obvious cases can be found on the quite global layer of the form of communication.

For instance, the personal blog *One Ideal Life* from 2011 is connected to the social network *Twitter*. The sidebar automatically displays the author’s latest tweets. This example shows how elements of one form of communication (*Twitter*) are “transplanted” onto another form of communication (*Blog*). The site’s author registered it as *personal weblog*, which underlines the assumption that the result is perceived as a blog containing Twitter elements, rather than as some new genre. This form of pattern embedding becomes widely spread from 2010 onwards.

4.4. Migration

Finally, I would like to sketch a process called migration. In this case, entire existing genres literally move into a new form of communication. The new medial environment then starts to exert its influence on the genre. Quite a number of Internet genres may have offline ancestors that migrated into the Internet and changed in the new environment. For *personal weblogs*, one possible ancestor can be assumed to be the offline *diary*. The migration from the form of communication *book* to an environment
created by blog software brought about changes such as link structure, comments or altered situational factors.

5. Conclusion

To sum up: In the present talk I have presented a diachronic approach to genre research on the internet. Even though the fluid surface of the WWW speaks another language, I assumed stable communicative patterns below the surface. A first glance at some preliminary results suggested how one Internet genre – the personal weblog – can be described diachronically. The observations also suggested sources of stability:

1. Migration provides for stability as many Internet genres are probably well-familiar from offline contexts.
2. Patterns widely known are combined in innovative ways.
3. Even on the WWW, patterns change only gradually following general trends.

Much further research needs to be carried out to elaborate the picture of generic stability and change on the Internet on all layers of the dynamic genre model. I think, after more than 20 years of WWW, it is high time we started telling its genre history, which is far from being chaotic. A diachronic approach as I have suggested it offers the exciting opportunity of observing Internet genres in the making.