“Please Check for Grammar.”: Code-Alternations in a Language Learning Blogging Community

Judith Buendgens-Kosten
Goethe Universität Frankfurt
Buendgens-kosten@em.uni-frankfurt.de

Abstract

This paper looks at code-alternations in the language-learning blogging community “Lang-8”. In this community, users write blog posts in their target language and receive feedback and corrections from native speakers. 116 blog posts with the target language English are analyzed in detail. Around 2/3 of those blog posts avoided all code-alternations. Among the remaining third, the most frequently observed type of code-alternation involved translation of part or all of the blog post (interlinear translation and en bloc translation), or switches motivated by lexical need (complex lexical gaps). Quoting, other sentence-level switches and word-level switches for meta-linguistic discussion were rare.

The author argues that technological affordances, different understandings of one's audience and of the language learning process, and, to a limited degree, community norms, shape code-alternations and their avoidance.

Keywords
code-alternation – code-switching – computer mediated communication (CMC) – second language acquisition (SLA) – blogging – computer-assisted language learning – computer-assisted language learning (CALL)

1 Introduction

Blogs, historically also called weblogs, are “frequently modified web pages in which dated entries are listed in reverse chronological sequence” (Herring, Scheidt, Bonus and Wright, 2004: 1), often with titles, tags, links, and embedded multimedia elements. Blogs are sometimes referred to as “online diaries”, but
so-called “journal-style” or “diary” blogs (Scheidt, 2006) constitute only a small portion of blogs overall. Blogs may be written by one (individual blog) or several (group blog) authors, referred to as bloggers. Blogging software frequently provides ways to connect with bloggers and to respond to individual blog posts, such as commenting, backtracking, or blogrolls, and for many bloggers the community of related bloggers, the blogosphere, is an important part of their blogging experience.

The topics covered by blogs are wide-ranging: some blogs focus on daily life experiences (journal-style blogs) or interesting links (filter blogs), while others, sometimes called knowledge blogs, are dedicated to specific topics such as sports, politics, or education. The fact that blogs are used to collect, order, present information, to report about experiences and to reflect upon them, and to discuss the content of blog posts through comments, make them suitable for personal knowledge management (Reinmann and Bianco, 2008). They are also increasingly used in formal educational settings. Major learning management systems (LMS), e.g. Moodle, provide in-built blogging functionalities for learners and teachers.

Blogs have also been used for language learning, usually in combination with traditional classroom teaching (e.g. Arslan and Şahin-Kızıl, 2010; Dippold, 2009; Ducate and Lomicka, 2008; Eastment, 2005; Fellner and Apple, 2006; Lee, 2010; Miceli et al., 2010; Mompean, 2010; Murray and Hourigan, 2008; Spanhel, 2007; Trajtemberg and Yiokoumetti, 2011; Wang, 2009; Zimmerer, 2009). In this discussion, the focus will be on settings where students blog, i.e. settings that are not limited to an “information hub” blog maintained by the teacher.

Blogs as technology enable a wide variety of scenarios for language learning classrooms. While the possibility to add multimedia elements such as video makes it possible to use blogs to practice speaking skills, most research focuses on aspects of developing written language skills. The existence of a (micro-) audience is often seen as encouraging accuracy in writing. In many contexts, the imagined audience (Ong, 1975) suffices: Students write for an abstract audience with which they do not interact, but which is seen as nonetheless motivating (Bündgens-Kosten and Brombach, 2013). In other cases, fellow learners function as audience, and may engage in explicit peer feedback and peer corrections (Dippold, 2009). The presence of an audience often serves as the basis for an authenticity claim (Buendgens-Kosten, 2013). Depending on the exact setting in which they are used, blogs can also be employed to develop fluency, as in Fellner and Apple 2006: “The instructors explained to the students that grammar and syntax were not being evaluated and that the only requirement for blog entries was that they write their honest opinions about the blog topics.” (Fellner and Apple, 2006: 17).
Beyond writing skills, metacognitive skills such as reflecting about the language learning process may also be encouraged via blogs (Mynard, 2008; Hourigan and Murray, 2010), as may be intercultural learning (Lee, 2009; Andrew, 2000; Ducate and Lomicka, 2005). Correspondingly, blogging-assisted language-learning (BALL) settings tend to be fairly diverse.

This paper looks at language use in a specific language-learning blogging community, Lang-8 (lang-8.com), which allows learners to blog in their target language(s). As a blogging community, Lang-8 not only serves as a platform for individual blogs, but also encourages connections and exchanges between different bloggers, including the correction of blog posts by native speakers. The focus will lie on bloggers with English as their target language, and their use of English and of other languages in their blog posts.

Unlike the scenarios discussed above, Lang-8 is not necessarily connected to a classroom or course context. Learners may use Lang-8 to supplement formal language study at school or university, but no such connection to formal language learning is required.

1.1 **Language Status Online**

In 1993, the New Yorker published a cartoon showing two dogs in front of a computer, with the caption stating: “On the Internet, nobody knows you’re a dog.” The cartoon has since developed into a meme of its own, adapted, modified, parodied in many different ways. The basic notion: Online, you can be whoever you want to be. Nobody knows who you “truly” are.

While in many online communities “faking”, i.e. assuming a new identity for purposes as wide-ranging as role-playing and scamming is possible and the more benign practices may actually be encouraged, this does not mean that features of identity as they are presented online are irrelevant. Kendall argues that while online communities such as MUDS provide individuals with more choice about how to represent themselves, “Participants’ offline identities underlie their online identities and conversations. (...) Gendered, classed and raced identities continue to have salience in online interactions, with power relations often operating in much the same ways as they do offline, even when participants understand that people’s online identities may differ from their offline identities.” (Kendall, 1998: 150). Nobody may know for sure if you are a dog, but if you present yourself as one, this will influence interactions online.

I argue that the same is true for speaker status such as “native speaker” or “language learner”, at least in the context of the Lang-8 blogging community. For users at Lang-8, it is technically possible to assume any native language they would like. But these choices have consequences.
Within the community, the status of members as native speaker of one,\(^1\) and non-native speaker of other language(s), is emphasized.\(^2\) This emphasis on speaker status, reflected by Lang-8’s software affordances (see section 3.1 and 3.2) and community norms (see section 3.3 and 3.4), makes Lang-8 an interesting context for studying language alternation. How do learners use their linguistic resources in such a setting? How do they construct themselves as learners/native speakers of a specific language? How do they use target language and native language for non-classroom language learning? Which effects do different understandings of the language-learning process or of one’s audience have on language choice?

2 Multilingual Practices in Blogs and Asynchronous Online Communication

Even though there has been a growing amount of research on code-alternation in CMC and on multilingual practices online more generally, a lot of this research is on synchronous communication, such as IRC (Paolillo, 2001; Siebenhaar, 2006), chat: including text message services such as ICQ and MSN (Lam, 2004; Decker and Vandekerckhove, 2012), MOOs (Kötter, 2003), or Twitter (Eleeta, 2012). Research on code-alternation in asynchronous communication is done less frequently. Seargant (2011) summarizes this as a perceived lack of code-alternations “on the public, asynchronous Web” (Seargant, 2011: 502). While some specific multilingual practices might be closely associated with informal, synchronous exchanges such as some chat formats, this does not imply that code-alternations are without relevance for asynchronous CMC, as this literature overview hopes to demonstrate, focusing on blogs as one representative of asynchronous CMC.

Please note that in the discussion of code-switching and code-alternation in (partially) English blogs, I take a very broad understanding of code-switching/code-alternation: Any use of English with another language will, at this point, be considered.

Auer defines language alternation as “the cover term of all instances of locally functional language usage of two languages in an interactional episode” (Auer, 1984: 7), which subsumes under it code-switching (“any language alternation at

---

1 The Lang-8 profile affordances do not allow entering multiple native languages.

2 Please note that I use “native speaker” and “non-native speaker” merely to describe categories meaningful within the community, i.e. as labels for describing a folk linguistic understanding of language status.
a certain point in conversation without a structurally determined (and therefore predictable) return into the first language" (Auer, 1984: 26)) and transfer ("any language alternation for a certain unit with a structurally provided point of return into the first language with that unit's completion" (Auer, 1984: 26)). In Gumperz' terminology, both conversational code-switching ("the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems" (Gumperz, 1982, 59)) and many forms of borrowing ("introduction of single words or short, frozen, idiomatic phrases from one variety into the other" (Gumperz, 1982, 66)) will be included, of course adapted to the written and asynchronous nature of blogging. For a description of the different types of code-alternation observed in the data, see Chapter 5.

Lienard and Penloup (2011) argue that code-switching in blog posts, especially code-switching to English, is closely associated with patterns of language use among young people: "We could find scores of examples because English is undoubtedly the language French youngsters will use the most to claim their group identity, whatever the communication situation or the media. But this is not exclusive to English." (Lienard and Penloup, 2011: 79). Šabec's (2009) description of Slovene-English code-switching on Slovene blogs might fit in here as well. The author does not provide information about the demographics of blog authors studied, but the features (beyond code-switching) of blogs discussed are closely associated with the more extreme – or, one may say, more playful – forms of youth language online.

Montes-Alcalá (2007) studied code-switching in bilingual Spanish-English blogs. Her paper discusses the explicit negotiation in blogs of the acceptability of blogging in more than one language. Statements made in blogs regarding multilingual blogging range from strict opposition (except to changing languages in between blog posts, as long as each language can be retrieved individually), to strong acceptance. Furthermore, she analyzed a corpus of 15 blog posts that show instances of code-switching to identify the functions these switches might have. Types of code-switching differed between blogs, but overall "Lexical items" (27.6% of all switches) and "Emphatic" (24.3%) were observed most frequently (Montes-Alcalá, 2007: 166).

How do adult bloggers for whom English is a foreign language that they learned through explicit instruction at school or university use English in their blogs? Yu Liu (2009) analyses English/Chinese code-switching in blogs of Chinese speakers studying abroad, providing a detailed analysis of grammatical features of intra-sentential code-switching in the context of Myer-Scotton’s matrix language hypothesis (Myer-Scotton, 1993). 37 of the 79 blog posts analyzed contained English-Chinese code-switching within the projection of complementizer, i.e. it must be assumed that the total number of blog posts
containing any form of code-switching is higher. Not all of these bloggers were actually living in an English-speaking (inner circle; Kachru, 1985) country at the time of blogging, though all lived in a foreign country where English played an important role in educational contexts.

Not all bloggers who blog in English live in an English speaking country, or in any other foreign country where they might use English as a lingua franca. One example is discussed by Leppänen (2007): the English blog of a Finnish teen, where Finnish names (places, family relations, etc.) are used, but no other code-alternations happen. Panke, Gaiser and Maaß (2012) describe another such case, i.e. German-speaking scientists who blogged about their research:

Dennoch werden teilweise Versuche unternommen, auch englischsprachige Blogeinträge zu verfassen. So bloggt eine befragte Person beispielsweise dann auf Englisch, wenn der Beitrag insbesondere für die internationale Leserschaft von Interesse sein könnte.

PANKE, GAISER AND MAASS, 2010: 10

[Nonetheless, there are occasional attempts at writing English blog posts. One interviewed person blogs in English if the contribution might be of special interest for an international audience.]

One reason to blog in English or in another non-native language is the opportunity to reach an international audience. For many scientists, though, this is a double-edged sword, since they need to signal membership in the local scientific community in addition to membership in an overarching international scientific community. Blogging in one language, with some blog posts in another language, can be understood as an attempt to signal this double membership.

This kind of multiple-language blogging, which uses different languages to reach different audiences, has received little scientific attention. Here, the individual blog post is monolingual, only the blog as an overarching structure demonstrates code-alternation.

Another type of blogging may be pointed out that is closely related to multilingual blogging: multi-variant blogging. Blog posts may systematically be written in multiple versions, usually for some educational purpose. The now defunct Hiragana Times blog (original URL http://www.hiraganatimes.com/hiragana_blog/, archived by http://archive.org/web/) is an example of this: Blog posts were provided in a combination of kanji, kana and Rōmaji, in pure hiragana, and in Rōmaji transcription. Another example is the “dual blogging”
category (http://www.forschungs-blog.de/category/dual/) on Forschungsblog.de, a German-language science blog. “Dual blogging” describes a practice of “bi-sociolectal” blogging. On the left side of the blog post science news are presented in a formal, academic register. On the right side, one can read a “blogaffine Perspektive des gleichen Themas, lebensnäher und (hoffentlich) unterhaltsam geschrieben, was sich vermutlich besser für die Social-Media-Welt eignet” [blog-suited perspective on the same topic, closer to life and hopefully written in an entertaining style, which probably makes them better suited for the social-media-world] (http://www.forschungs-blog.de/dual/).

As discussed above, blogging is a popular way to integrate computer-assisted language learning (CALL) into face-to-face classroom settings. A number of studies report on the opportunities but also the difficulties involved in blogging in the SLA classroom, few of these studies take up the topic of using both the target language and the source language(s) in student blogs, though, probably because a monolingual (target language) norm is implied in the pedagogic design (Liegenhausen, 1991: 61). The following papers are among the few exceptions.

Exchange blogs, either based on travel experiences (Ducate and Lomicka, 2005), or on intercultural project work (Lee, 2009), sometimes actively embrace multilingual practices. The principle of one text=one language is usually maintained in these blogs, at least on the level of individual posts, but the blog as a whole, i.e. as a collection of blog posts and blog comments, can combine several languages. For example, in the blog project that Ducate and Lomicka (2005) describe, in a shared group blog US university students on an exchange program in France blogged in English, while French students blogged in French. In the project described by Lee (2009), American students commented on English blog posts written by Spanish students in Spanish, creating an overall mix of languages within the blog, while keeping individual blog posts or comments monolingual.

Suggestions that the intentional switching between languages (by the same writer) may be beneficial, are rare in BALL contexts. Murray and Hourigan (2008: 94) pose an exception here. They suggest allowing use of native language for reflection as one option for encouraging and simplifying student reflection through blogs. This would result in content posts written in the L2 and reflective posts written in the L1, published in the same blog.

2.4 Summary
So far, we have discussed code-switching or code-alternations in a number of different blogging scenarios, from a teenager’s personal blog to the blog post addressed to an international community of scholars. Switching to or
borrowing from English in blogs is often observed as part of youth language, i.e. the medium used by young people for group communication in specific settings (Neuland, 2007: 13) Outside of youth language, its acceptance in blogs differs widely. Switching languages between blog posts seems to be more acceptable than writing one blog post in more than one language. This practice, while frowned upon by some, can still be observed in some blogs, depending very much on the specific type of blog/blogger.

In all of this, however, one should not forget one more “type” of code-alternation: The avoidance of code-alternation. This seems, by far, the most common choice made by individuals with the linguistic skills to blog multilingually, on the level of individual blog posts (intra-post alternations) or of whole blogs (inter-posts alternations).

This study focuses on blogging for language learning in non-formal contexts, i.e. independent of classroom learning. None of the studies discussed above have looked at a similar context – though both informal contexts and language learning contexts have received attention. In the following section, I will introduce the Lang-8 community in more detail, in order to provide sufficient background for the discussion on (intra-post) code-alternation behavior.

3 Lang-8 and Blogging in the Target Language

At Lang-8 (Bündgens-Kosten, 2011), language learners write blog posts in their target language, which are corrected and commented on by native speakers of this language. Lang-8 is an example of reciprocal language teaching (Cook, 1989), but does not focus on 1:1 relationships like language tandems do. While each person at Lang-8 takes up learner and teacher roles, these can be in relationship to different users, and they might be very temporary, even limited to only one exchange. For example, a native speaker of Japanese learning English may receive corrections and feedback from a native speaker of English learning German, whose blog posts are corrected by a native speaker of German who is learning Japanese (see Image 1). Currently, Lang-8 provides language tags (blog post language identification) for 80 languages, from Afrikaans to Yiddish, plus a website interface in 18 different languages/writing systems.

According to Mizumoto et al., Lang-8 has more than 200 000 users, 142,311 of which are learning English, the most frequently studied language on Lang-8 (followed by Japanese with 75,162 learners, French with 16,711 learners, Mandarin with 15,434 learners, Spanish with 13,380 learners, Korean with 12,401 learners and German with 11,104 learners (Mizumoto et al., 2011: 148)).
Lang-8, as a community of bloggers, is more than a mere collection of individual blogs. It integrates the blogging experience with elements of social networking (Rezaei, 2010; Harrison and Thomas, 2010).

3.1 **Lang-8 as Blogging and Social Networking**

Lang-8 contains a number of features typical of blogging more generally. Each blog consists of a short user profile, blog posts, and additional information about the user such as number of posts written, number of posts corrected, testimonials and contact information. Each blog post consists of the blog post text, a title, and metadata such as tags (keywords), a language tag, a timestamp, and author information. Multimedia content such as audio, video and images might be embedded into blog posts, as well as links to other websites. Simple formatting options are available. Older posts can be accessed through an archive (list of older blog posts) and a tag lists (list of key words and associated blog posts). Readers can respond to each blog post through comments and corrections, the second of which is not typical for blogging more generally. RSS feeds are provided, but users can also “befriend” a blogger to be made aware of new blog posts.

The asynchronous nature of blogging means that bloggers have time to carefully craft and edit their blog posts, and to use a wide range of online tools such as online dictionaries or inflection tables while writing, in addition to any printed resources that may be used. Lee (2009) argues that “(O)ne of the major benefits of using asynchronous CMC is the opportunity to focus on form” (Lee, 2009: 430). Whether this focus on form is desirable for language learning through blogs, though, or whether, as e.g. Rivens Mompean argues, “Students need to be somehow deconditioned to focus less on form, as they have been taught during their whole learning process, and to dare to communicate focusing more on fluency and less on accuracy “ (Rivens Mompean, 2010: 391) remains debatable.

The fact that texts in blog posts do not need to stand on their own has an additional effect here. Bloggers can embed images and other forms of media to...
illustrate their ideas, and use hyperlinks to refer to external texts and media. Some simple blog posts at Lang-8 may consist only of a picture and a descriptive sentence, while others may combine embedded videos, own texts, quoted text, and links to external sources. The multi-media and hyperlinking features of blogging thus support both beginners and more advanced language learners.

For the purpose of this paper, we will consider the individual blog post as the unit of blogging. Each blog post might be fairly long, longer than e.g. status updates, tweets or typical chat messages. While the asynchronous nature of blogging and the options for including a wide range of material for clarification may lower the cognitive and linguistic requirements of writing for non-native writers, the expected length of texts (which naturally differs between subgenres and communities) can constitute an additional challenge.

As stated above, Lang-8 combines features of blogging with features of social networking. The social networking features primarily support the establishment and maintenance of (direct or indirect) reciprocity. They help users to identify bloggers writing in their native language who require assistance with their language learning, and provide features that reflect everybody's engagement in supporting others in their learning process (a practice associated with signaling online reputation). “Befriending”, i.e. the formation of stable learning-support dyads is also encouraged. Similar to social networks, bloggers can decide with whom to share their blogs by modifying the privacy settings for each blog post: visible to nobody, only their friends, every user, everybody.

3.2 Lang-8 as Language Learning Practice
In addition to the affordances discussed above which reflect typical features of blogs and of social networks, Lang-8 also provides features very specific to language learning. Lang-8 is about language in the sense that all activities at Lang-8 are centered around language learning. This is also reflected in the structure of the website, which is centered around the language blog posts are written in. You can add content tags to your blog post if you like, but the language tag is required. If you wish, you can add information about your gender, age, location and hobbies to your profile – but native language and target language(s) are mandatory information available to all users. On your homepage (the “dashboard”-like view you get upon login) under the most current posts written by your friends, there are lists of posts in your native language (“new posts needing correction”) and lists of posts in the language(s) you are learning. While these affordances do not enforce specific behaviors, they provide what van Leeuwen calls “easy-to-choose default options” (van Leeuwen,
A comparison between Lang-8 and other blogging communities, e.g. livejournal (Kendall, 2007; Campbell, 2004), shows how drastic the differences are even on the basis of the user interface alone. One may even say that language and linguistic correctness are the social object of Lang-8, i.e. what “mediate(s) the ties between people”, or constitutes the “reason why people affiliate with each specific other and not just anyone” (Engeström, 2005): Two bloggers connect because one blogger is a native speaker of the language the other blogger is currently learning and can provide corrections, not because they share, e.g., the same hometown or hobby.

In such a context, linguistic choices, language choice and code-switching among them have a much higher salience than in other contexts. Gumperz’s observation, that “participants immersed in the interaction itself are often quite unaware which code is used at any one time” and that “(T)[h]eir main concern is with the communicative effect of what they are saying.” (Gumperz, 1982: 61) may not hold here.

The “communicative primacy”, i.e. the observation that “(I)f attitudinal factors (i.e., the sorts of social prejudices which create linguistic stereotypes) do not intervene to foreground some structural element, such elements appear to be overwhelmingly subservient to communicative function” (Niedzielski and Preston, 2003: 16ff) is regularly violated at Lang-8.

Also, I suggest that a monolingual ideal is encoded in the technical make-up of the community itself. At Lang-8, you are required to assign a language tag to each blog post before publishing it online – but each post can only be assigned one language tag. While it is of course possible to write a blog post in multiple languages, it cannot be marked as such. This might be interpreted as a “hard-coded” norm: While languages may vary throughout the blog, it is assumed that a blog post as such is monolingual.

### 3.3 Role of Correction and Commenting

At Lang-8, your native language - as defined by your profile setting - determines how texts are presented to you. Posts in your native language are not offered for your reading pleasure; they are presented as texts to be corrected, while those written in your target language(s) are offered as input or reading material.

While Lang-8 allows readers to comment on blog posts, its correction functionality is visually more dominant: If you wish to correct a blog post, a copy of every sentence written by the original poster is provided for your convenience, and editing marks such as strikethrough or red ink are suggested as formatting options. Bloggers often explicitly ask for corrections and feedback on their text, as was the case in the blog post quoted in the title of this paper: “Please check for grammar.” Receiving corrections is seen as desirable. As far as rating
the quality of blog posts, a traditional feature of social networks, Lang-8 users do not “like” a post, but give a “native speaker nod”. Through all these design features and affordances, the role of the native speaker and the role of the language learner are stressed and defined in a specific way. You are not just a person communicating in language x, you are a native speaker of y and a learner of z.

Burt (1992) studied the code-switching behavior of dyads where one person was a native speaker of the language the other person was studying, and vice versa. Though her scenario focused on face-to-face exchanges, the language-status-combination is similar to what happens at Lang-8. Burt describes this setting as a rights and obligations (RO) set (Myers Scotton, 1983):

Regardless of whether language A or language B is chosen, one of the dyad will be cast in the role of native speaker, while the other will take on the role of non-native speaker. The role of native speaker entails the right of acting as authority in deciding the right way to speak the chosen language, and the obligation of offering help in the form of instruction to the non-native speaker. The role of non-native speaker entails the right to help, patience, simplified speech or perhaps overt instruction from the native speaker, and the obligation to submit to correction from the native speaker if this is forthcoming.

Burt, 1992: 171

At Lang-8, one does not choose language depending on one’s target audience; instead, one chooses one’s audience based on the language one wishes to practice. Lang-8 is a site of constant convergence: You log onto Lang-8 to write in a language that is not your native language, and to read texts written in your native language by people currently learning it.

Burt, in her discussion of face-to-face language learning dyads, argues that “incipient bilinguals in conversation with each other have no such community norms to fall back on - or even to acquire” and that “successful language learners will manage to create such norms in the process of establishing relationships with speakers of the language they are learning.” (Burt, 1992: 184)

At Lang-8, though, exchanges can be fleeting. The person correcting your blog post today might never read another of your blog posts again – especially if you are writing in one of the frequently studied languages such as English or Japanese. If you do not establish more permanent dyadic relationships through other means (such as “friending” another blogger), it may not be possible to negotiate such rules or to learn the expectations of the individuals reading your blog posts. On the other hand, Lang-8 as a whole, that is the Lang-8
community in the broadest possible sense, may potentially provide certain norms, beyond those implied by the affordances discussed above.

3.4 **Lang-8 as a Community**

Lang-8 is clearly a community in the broadest sense of the term, but is it a (virtual) community in a stricter sense of the term? According to Herring’s framework of computer-mediated discourse analysis (Herring, 2004: 355f), virtual communities have six features: (a) active, self-sustaining participation around a core of regular participants, (b) evidence of shared history, culture, norms and values, (c) solidarity and support, as evidenced by, for instance, humor, positive politeness, and reciprocity, (d) criticism, conflict, and the emergence of means of conflict resolution, (e) self-awareness of the group as an entity that is distinct from other groups, and (f) emergence of roles, rituals, and hierarchies. Lang-8’s ‘track record’ here is mixed. Lang-8 has “active, self-sustaining participation; a core of regular participants”. Its “shared history, culture, norms, and values” is demonstrated – beyond the mere technical feature of availability of older posts - by a shared understanding of the purpose of the community (language learning) and a shared appreciation for linguistic feedback demonstrated in posts and comments. At the same time, there are no indicators of a developing “lingo” (abbreviations, jargon) or of insider jokes/memes, and explicit discussion of norms and values is rare or non-existent. Reciprocal corrections, use of thanking buttons, requests for favors, congratulations on good blog posts (via comments or “native speaker nods”), as well as the use of humor indicate “solidarity, support, reciprocity”. The other side of the coin, “criticism, conflict, means of conflict resolution”, is less visible. There are very few indicators for conflict; criticism takes place mostly on a meta-linguistic level, e.g. through formal means of correcting language mistakes. The “self-awareness of group as an entity distinct from other groups” is also only marginally developed. Occasionally, bloggers discuss how they discovered Lang-8, what expectations they associate with blogging at Lang-8, or which plans they have made for studying languages at Lang-8, but many of these posts are introductions to Lang-8, i.e. by as-yet peripheral members. The “emergence of roles, hierarchy, governance, rituals”, on the other hand, is very evident: the role distinction between native speaker and learner permeates the whole community (Herring, 2004: 355).

That Lang-8 does not achieve a clear verdict here has to do with the structure of interactions springing from its basic design as a language learning community in which learners communicate with native speakers: At Lang-8, two members may not share a single language. A native speaker of Danish learning English and a native speaker of Japanese learning Mandarin may not be able to communicate directly. In this way, the basic communicative structure
at Lang-8 differs dramatically from e.g. diaspora communities online (Androutsopoulos, 2006) or other types of bilingual communities online where several languages are used, but where each person (ideally) understands all languages used. Generally speaking, most communication seems to happen between native speakers and learners of a language, while exchanges between different learners of a language and between different native speakers of a language are possible, but less frequent.

Lang-8 also has an in-built “group” function, which is mostly in English as a lingua franca. Here, people who are able to use English, independent of their status as current learner, non-native speaker, or native speaker, meet and discuss. This part of Lang-8, though, is very infrequently used.

A consequence of these two observations is that the development of norms that transcend all language barriers cannot necessarily be expected. The frequently learned languages, such as English or Japanese, will have a certain unifying effect, but newly developing standards may still not reach every participant within these sub-communities. Also, membership at Lang-8 is often rather transient. It takes time to get to know a community and its habits, and for short-term members, this knowledge may not be available before they leave the community again, especially since the kind of texts written and the lack of an overarching topic (beyond an interest in the medium of exchanges, that is: language) make extensive lurking as a preparatory step to blogging unlikely.

Overall, one may prefer to use Gee’s notion of “affinity group”, a term used to describe ‘spaces’ where people interact, which may or may not be communities (Gee, 2004), to describe Lang-8.

This mixed “track record” must also be expected to influence code-switching behavior. Concerning conversational code-switching, Gumperz argues:

> Since bilingual usage rules must be learned by living in a group, ability to speak appropriately is a strong indication of shared background assumptions. Bilinguals, in fact, ordinarily do not use code switching styles in their contact with other bilinguals before they know something about the listener’s background and attitudes. To do otherwise would be to risk serious misunderstanding.

GUMPERZ, 1982: 69

When such a set of “shared background assumptions” is missing, code-switching might be discouraged. So, just as there is a low probability of two users developing their own rules for their exchange, there are also only limited community norms to guide them. This might strengthen the effect of community norms “hard-coded” into the software in the form of affordances.
4 Data Collection

A small corpus of blog posts was created. The sampling procedure consisted of collecting all blog posts (a) with the language tag “English”, i.e. the language practiced most frequently at Lang-8 (Mizumoto et al., 2011), (b) that were written in a specific time frame – 8 one-hour blocks distributed over two non-adjunct days (4 blocks evenly spaced throughout each day), and (c) which had the privacy setting “share with all internet users”. After data collection, blog posts without at least some identifiable English text were excluded. This resulted in a corpus comprising 116 blog posts by 115 bloggers.

Table 1 shows the self-identified native language of bloggers, as documented in user profiles.

Other demographic data is available only for parts of the sample, since information about, e.g., gender and age are not mandatory parts of a user’s profile.

5 Data Analysis

Blogging comprises a number of different practices. Many practices take place on the level of the individual blog post, i.e. writing blog posts, layouting them, embedding media, hyperlinking to texts and media, or adding metadata such as tags to them. Blogging, though, also includes practices that go beyond the individual blog posts, such as maintaining a user profile, choosing a design for your blog (not possible at Lang-8, but customary on other blogs), and filling different “slots” in a design with texts and images. Other practices central to the community of bloggers (“blogosphere”) surround blogs, but are not usually subsumed under the term blogging, such as reading blog posts of other bloggers or writing comments.

As a consequence of these different practices, each blog contains a multitude of texts, from blog posts over blog comments to short snippets “hidden” in the layout of a blog, such as lists, names, or profile information, some of which may be auto-generated (such as post archives), and some of which may have been written by the blogger (such as some of the profile texts). The focus in this paper, though, will be exclusively on the actual blog posts,

---

3 An alternative might have been e.g. Nishimura’s approach of labeling cmc as either “mostly language a” or “mostly language b”. It is not the amount of English included that leads to the identification of the text as English, but the intention of the author as manifested in the language tag.
including their titles. Furthermore, the analysis will look at blogs at the level of *individual* blog posts. This means that alternations of language between posts (e.g., in the case of learners of multiple languages) will not be considered here.

In a first step, all blog posts that contained evidence of code alternations of any type were identified. In doing so, I followed a very broad understanding of the term *code-alternation*, in the sense of Auer’s “cover term of all instances of *locally functional language usage of two languages in an interactional episode*” (Auer, 1984: 7), taking each blogpost as (part of) an interactional episode. Proper nouns, specifically names of people and places, movies, TV shows and characters, as well as morphemes that may be considered part of these (Japanese -kawa/-gawa as part of a river’s name, or Japanese -kun as part of a person’s name) were not considered, on their own, as evidence of code-alternation, even if given in a non-Latin script, unless their literal meaning was explicitly discussed or an attempt at explicit translation was undertaken. In the case of Japanese, words that have gained popularity in the J-pop or Anime

<table>
<thead>
<tr>
<th>L1</th>
<th>Absolute numbers</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Cantonese</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Catalan#</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Indonesian*</td>
<td>2</td>
<td>1.7%</td>
</tr>
<tr>
<td>Japanese</td>
<td>79</td>
<td>68.1%</td>
</tr>
<tr>
<td>Korean</td>
<td>9</td>
<td>7.8%</td>
</tr>
<tr>
<td>Malaysian</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Mandarin</td>
<td>4</td>
<td>3.5%</td>
</tr>
<tr>
<td>Polish</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Portuguese (Brazil)*</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Russian</td>
<td>7</td>
<td>6.0%</td>
</tr>
<tr>
<td>Tagalog</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Thai</td>
<td>1</td>
<td>0.86%</td>
</tr>
<tr>
<td>Traditional Chinese</td>
<td>6</td>
<td>5.2%</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>1</td>
<td>0.86%</td>
</tr>
</tbody>
</table>

* language as referred to by speaker
# Since other profile data are implausible, this may be an incorrect L1 as well.
subcultures (e.g., *enka*, *manga*, *j-pop*), as evidenced by inclusion in a popular English-language manga encyclopedia, or that have become established loan words of English outside of these subcultures, usually technical terms (e.g., *haiku*, *kanji*, *hiragana*) were also not counted as sufficient evidence for code-alternation. Specialized religious terminology, foods, literary terminology, names of holidays, etc. (e.g., Showa no Hi, だまこ鍋) were counted as examples of code-alternations, though.

In a second step, different types of code-alternation were identified by open coding. In addition to assumed functions of code-alternations, particular features of the layout or arrangement of the languages within the blog posts were taken into account. All coding was done using the qualitative data analysis software MAXQDA (www.maxqda.com).

Table 2 provides an overview over the codes and code-groups developed via open coding, and the frequency with which code/code group was used. The different coding categories will be discussed in detail in the following sections. Note that blog posts can be coded with multiple codes, e.g. as containing both (at least) one instance of quoting and (at least) one instance of complex lexical gaps. Two codes pose exceptions here: the code “avoidance” is only used if none of the other codes can be applied, and “excluded from analysis” means that no further coding has been attempted. Any blog post that contains at least one sentence-level code was counted as “all sentence-level” in the table, any blog post that had any word-level code was counted among “all word-level”.

**Table 2  Frequency of codes (N = 116 blog posts)**

<table>
<thead>
<tr>
<th>Code</th>
<th>Parent code</th>
<th>Number of documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>avoidance</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>all sentence-level</td>
<td>sentence-level</td>
<td>22</td>
</tr>
<tr>
<td>interlinear translation</td>
<td>sentence-level</td>
<td>16</td>
</tr>
<tr>
<td>en bloc translation</td>
<td>sentence-level</td>
<td>3</td>
</tr>
<tr>
<td>quoting</td>
<td>sentence-level</td>
<td>2</td>
</tr>
<tr>
<td>other sentence-level</td>
<td>sentence-level</td>
<td>4</td>
</tr>
<tr>
<td>all word-level</td>
<td>word-level</td>
<td>17</td>
</tr>
<tr>
<td>complex lexical gaps</td>
<td>word-level</td>
<td>14</td>
</tr>
<tr>
<td>metalinguistic discussion</td>
<td>word-level</td>
<td>3</td>
</tr>
<tr>
<td>excluded from analysis</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
6 Results

The default behavior was not to alternate languages. 79 posts (68.1%) contained no evidence of a second language, beyond, of course, unintentional transfer. 37 (31.9%) posts contained some evidence of code-switching or code-alternation in the widest possible sense. In the following section, I will discuss both code-alternation avoidance as well as the types of code-alternation observed, both on the sentence level and on the word-level. Examples will be presented throughout.

6.1 Code-Alternation Avoidance

Code-alternation avoidance hardly requires an example. Nonetheless, it is important to point out that the monolingual English posts, while sharing code-alternation avoidance, differed greatly. Images 2, 3 and 4 demonstrate the diversity found among monolingual posts in this sample. Some blog posts were very short, written in simple sentences, and deviated significantly from native speaker norms regarding orthography, grammar, and/or lexicon (Image 2). Some of these resembled classroom tasks, or even grammar drills (Image 3), others were reports about daily life and culture. Other blog posts, written by authors who aimed to entertain or to educate their readers, demonstrated more advanced English skills. Image 4 shows a blog post which is not only more linguistically complex than that shown in Image 2, but also of a content matter more likely to appeal to a reader. Code-alternation avoidance was not, therefore, an indicator of advanced language skill, nor of the lack thereof.

6.2 Sentence-Level Code-Alternations

Sentence-level code-alternations were observed in 22 blog posts (19.0%) (three blog posts included several types). Most of these (19 blog posts, or 16.4%) were examples of interlinear or en bloc translations, with only few instances of quoting (2 blog posts, or 1.7%) or other (4 blog posts, or 3.5%) switches.

6.2.1 Interlinear and En Bloc Translation

Intersentential code-alternations often took the form of a kind of gloss or translation, with one sentence in L2 followed by the same sentence in L1 (interlinear translation, see Image 5), or with a longer block of text in L2 followed by the translation of this block in L1 (en bloc translation, see Image 6). Interlinear translations were more frequent (16 instances, or 13.8%) than en bloc translations (3 instances, or 2.6%), though the categories sometimes overlapped, e.g. when a blog post contained only one sentence, or when the blogger...
switched from translating block-by-block to translating sentence-by-sentence (possibly viewed as very short blocks). In addition, in two out of the three en bloc translations, the translation was from an English original (e.g. a newspaper article as in Image 7) into Japanese, while in all cases of interlinear translations both language versions were composed by the blogger. Interlinear translation might, therefore, be considered the more typical type of translation work observable at Lang-8.

Often, only the main or central part of the blog posts was translated, with a monolingual English introduction or conclusion. Image 5 provides an example of this, and shows how the English conclusion has been set apart from the rest of the blog post with a line of hyphens, suggesting a certain discontinuity between the part above and the part below the line. In one instance, complex combinations of L1 only, L2 only, and translation work could be observed (see Image 7). Some bloggers also included a title in two languages, or provided a translation of the title in the first line of the blog post, while others used a monolingual title.
6.2.2 Quoting

In the section above, Image 7 presented an example of complex translation work. Here, some of the English text was a verbatim quote from a newspaper article that is being discussed in English and has been translated into Japanese.

Quotes from the L1 in a mostly L2 text could also be observed in a few (2, or 1.7%) cases. In these instances, no translations were provided. Image 8 shows one of these. Here, Japanese haiku are included in their original language, without translation (the other code-alternation observable here will be taken up in the next section).

6.2.3 Other Forms of Sentence-Level Code-Alternation

Very rarely (4 instances, or 3.5%) could sentence-level code-switches be observed that did not fall into one of the above categories. In three of those, a short Japanese-only section was attached to the end of the blog post. Image 9 presents a typical example.

In Image 9, we have a Japanese end section, in which the author summarizes the ideas discussed in the blog and further elaborates on them: “The number of suicides during the war was shocking, even though it has now diminished. What kind of suffering was more painful, the suffering now or the suffering during the war?” The text-final position seems to be typical for this type of
code-alternation. Note how the Japanese section is only set apart by a new paragraph and empty line, unlike in Image 5, where a stronger visual demarcation between the interlinear translation and the monolingual English conclusion section was provided.

As stated above, an L1-only section could be found at the end of three of these blog posts. The fourth case was a blog post where the blogger alternated between two L2s, and started the blog posts off with a short, non-English L2 introduction. In this example, the switches to the second L2 were in the form of short, very simple sentences or sentence fragments, which might be considered emblematic switches (Legenhausen, 1991: 63). All this makes this blog post rather unusual.
Unlike the other blog posts discussed above, these blog posts are most meaningful to those who speak both English and the L1 or the second L2 used. A reader who understands only some of the languages used may not be able to follow the content of the whole blogpost. In this sense, they differ from “pure” interlinear and en bloc translations, which are accessible even if only one of the languages included is understood. This kind of language use, though, seems to be exceptional at Lang-8, as the low proportion in this corpus attests.

6.3 Word-Level Code-Alternations
When switches happened on the word-level (in a total of 17 blog posts, or 14.7%), this was not usually due to a learner’s lexical need, in the sense of mere lack of vocabulary knowledge, but instead due to very culture-specific terms for which no ready-made target language term exists (true lacunae) or can easily be found using general-use dictionaries (examples: names of specific rules for haiku writing, names of deities, etc.). I refer to these as “complex lexical gaps” (14 blog posts, or 12.1%). Other switches occurred in the process of asking metalinguistic questions (4 blog posts, or 3.5%).

6.3.1 Complex Lexical Gaps
As stated above, 14 blog posts contained examples of what I call “complex lexical gaps”. These were instances where a term that is not easily
translatable was included in the L1, often with further explanations. Image 10 demonstrates how these gaps were turned from a mere “lack of lexicon” into an opportunity for teaching language and/or culture. Another good example is included in Image 8, where the term “ginko” was not only introduced, but the reader was actually made aware of a homophone it can easily be confused with.

Contrast this with Image 11. Here, a term many readers may be unacquainted with, *Puri* (short for *purikura*, decorated photo stamps), is used, but not explained, nor is its status as a nonce borrowing signaled in any way (e.g. by including it in katakana script or using an italics font). Perhaps due to lack of awareness that this term may pose problems for native speakers of English, or because a specific audience (such as Japanese learners of English, or native speakers of English living in Japan) was addressed, this “teaching opportunity” is not used. That the post was nonetheless successful is evidenced by the comment in which a native speaker of English, who has indicated his location on his profile as “Japan” and reports previous activities as an English teacher in Japan in his profile text, suggests replacing *Puri* by *purikura*.

**IMAGE 10**  
Blog post by たくや: “Drinking Party”

**IMAGE 11**  
Blog post by NaKo: “Shop Till I Drop in Nippori”
6.3.2 Metalinguistic Discussion

Unsurprisingly, individuals used elements from their L1, or from another language they spoke, in metalinguistic questions, as illustrated by images 12 and 13. This category does not encompass all metalinguistic discussion, as some form of metalinguistic discussion also takes place in the case of complex lexical gaps, discussed above. Instead, it was only applied in those cases in which a word was used merely to ask a question about it. This could be observed in three blog posts (or 2.6%).

Note how both blog posts are used to ask questions that could not easily be answered by referring to a dictionary or other online tools (as the author of the blog post included in Image 12 clearly had done before posting). Also note how the author of the second blog post, Image 13, avoided code-alternation even for metalinguistic discussion in most instances by suggesting multiple near-synonyms, and only switched to Korean when this strategy failed, setting the Korean off in a footnote, creating a separate space for it, distinct from the rest of the text.

7 Discussion

This paper started out with the intention to study code alternation and how it reflects the specific communicative context and the differing perceptions of the community and the communicative setting. What do the code-alternation patterns observed tell us about language learning at Lang-8?

*Writing a monolingual text is not an “accident”.* The high number of blog posts without any form of code-alternation serves as evidence for the existence of a monolingual ideal on the text level, i.e. the ideal that one text should be...
written in one language. In many ways, this monolingual ideal is hard-coded into software affordances such as the language tag function.

In this context, code-alternations must be considered conscious choices. Importantly, code alternations do not indicate missing language skills. While many bloggers at Lang-8 have very limited language skills, these are usually reflected by simple sentence structures and short texts. The availability of language resources such as online dictionaries, as well as the time available to carefully plan, draft, and correct any text, mean that few code alternations are created out of linguistic necessity.

The category “complex lexical gaps” only superficially poses an exception here. These gaps are often lacunae, untranslatable or at least untranslatable outside of a specialist’s lexicon. Bloggers use these lacunae in their texts as opportunities to teach about language and culture. Instead of demonstrating a language deficit in their target language, they demonstrate expertise in their native language(s) and culture(s).

The dominance of translation as the most frequently used form of sentence-level code-alternation cannot be understood without reference to the dedicated purpose of the community: Learning languages. Many bloggers will be acquainted with grammar-translation models of language learning and used to translation tasks being done as part of the language learning process. Learning with parallel texts – a technique currently popular in online language learning forums and blogs - also may play a role for some learners. In a grammar-translation context, the learner produces both language versions as part of his/her own learning process, and to facilitate correction of his/her text. In a parallel texts context, the blogger provides the version in his/her native language as a service to those learning it, so that they, while correcting his/her L2 texts, can learn his/her L1 language.

Above, we have discussed Gumperz’s observation that the “ability to speak appropriately is a strong indication of shared background assumptions” (Gumperz, 1982: 69), and why it may be difficult to develop those shared background assumptions at Lang-8. This does not mean that bloggers do not have any vision of their audience. They are – necessarily – writing for an imagined or fictionalized audience (Ong, 1975), and, possibly in addition to that, for a concrete sub-audience (e.g. specific friends or a classroom teacher who has given an assignment). The context of blogging at Lang-8 will influence one’s understanding of one’s (imagined) audience as well. Lang-8 emphasizes corrections over comments, and language learning over exchange of ideas. Still, there might be a certain variance in how bloggers envision their audience. Some texts, most conspicuously those that contain grammar drills, but also, to a lesser degree, those that contain school-essay type texts, may be written for
an audience viewed primarily as a teacher (reader-as-teacher), while those that set out to teach content matter, language and culture, or that try to establish rapport through diary-type entries, may not be written for such a reader-as-teacher audience. This, of course, is not a clear-cut distinction, and bloggers' understanding of their audience may change over time.

8 Conclusion

The patterns of language choice observed are probably very specific to Lang-8, and may not be generalized to other language learning communities or affinity groups. I would suggest, though, that those elements found here to have an impact on language choice might also be at play in quite distinct communities/affinity groups. Those are, specifically, that

- technological affordances influence language choice,
- bloggers' understanding of the language learning process influences language choice,
- bloggers' assumptions about their audience influence language choice,
- and that, for long-term members or (ex-)lurkers, some (probably language-specific) norms concerning language choice may develop.

These elements, of course, interact with each other. The way bloggers use affordances can impact design decisions, i.e. indirectly create new affordances.

Since this study was completed, Lang-8 has introduced a new feature called “Write in your native language”, which capitalizes on and formalizes a specific form of code-alternation: one text field is presented for the target language text, while another text field is presented for the native language version of this text. It must be assumed that the introduction of this feature is related to the fact that a certain number of users were actively switching between languages in the form of interlinear or en bloc translations. Code-alternation was introduced, though, in such a way that it does not undermine the monolingual ideal underlying Lang-8. The new function creates a second, separate field for blogging in one’s native language that is visually separate from the target language section. Even though one way of using two codes in one blog post – similar to en bloc code-alternations – is now supported by the affordances of Lang-8, “mixing” languages – be it for lexical gaps, for creating rapport, for quoting, or for acknowledging a community language, still is not a central practice in this community. The new function provides each language with its “separate, but equal” place.
References


