Learning a Language for Free While Translating the Web.
Does Duolingo Work?

Ignacio Garcia

1 School of Humanities and Communication Arts, University of Western Sydney, Australia

Correspondence: Ignacio Garcia, School of Humanities and Communication Arts, University of Western Sydney, Locked Bag 1797, Penrith, NSW, 2751, Australia. Tel: 2-9772-6540. E-mail: i.garcia@uws.edu.au

Received: November 24, 2012  Accepted: December 24, 2012  Online Published: January 25, 2013
doi:10.5539/ijel.v3n1p19  URL: http://dx.doi.org/10.5539/ijel.v3n1p19

Abstract
Duolingo, a free online language learning site, has as its mission to help users to learn a language while simultaneously using their learning exercises to translate the web. Language is learned through translation with, according to developers, Duolingo being as effective as any of the leading language learning software. For translating the web, machine translation is not good enough and relying only on professional translators, far too expensive. Duolingo, we are told, offers a third way, with translation as a by-product of its language learning. Translation which will be, if as promised, almost as cheap as if done by machines and almost as good as if by professionals. Launched in June 2012, Duolingo boasts already at the time of writing 300,000 active language learners ready for the task. This article independently assesses the extent to which Duolingo, at its current stage of development, meets those expectations.

Keywords: online language learning, translation, machine translation, crowdsourcing, Duolingo

1. Introduction
Those interested in learning a language (English<>Spanish, English>French, English>German or English>Portuguese at the time of writing – the language of the learner matters since teaching is based on the translation method) should consider Duolingo. It is free, it will remain free, they promise, and it claims users can learn as effectively as if with any of the other leading language learning software.

Researchers and teachers involved in language learning should pay attention to Duolingo too. For centuries, translation was the default methodology for formal foreign language learning. Displaced by the communicative approach in the nineteen sixties, it has hardly been used for that purpose since (Richards & Rodgers, 2001). Translation comes back in force with Duolingo, reclaiming the centre stage once had.

Translators and researchers on translation should also look at Duolingo with interest. For translating the web, machine translation is not good enough and relying only on professional translators, far too expensive (Esselink, 2001). Duolingo offers a third way, with translation as by-product of language learning. Translation which will be, if as promised, almost as cheap as if done by machines and almost as good as if by professionals.

Behind Duolingo there is a team headed by Luis von Ahn, a renowned computer scientist at Carnegie Mellon University. If this wasn’t the case, Duolingo’s ambitious claims would have likely been disregarded by the learned community, but with von Ahn at the helm, the person behind such successful ventures as CAPTCHA and reCAPTCHA, it must be taken seriously.

Learners have done so. Launched in June 2012, Duolingo boasts at the time of writing over 300,000 active users, activity measured as a daily average of 30 minutes involvement (Farber, 2012). Venture capital is betting on its success, with some 15 million already invested (Farr, 2012). A great deal of attention has been paid to it (see for example Siegler, 2011 and Segumpta, 2012). Yet, all the information we have on it can be condensed on a few paragraphs of the von Ahn’s TED talk in which he announced it in April 2011 (Ahn, 2011). This talk remains the best summary on how the system works and what its aims are. The present article independently assesses the extent to which Duolingo, at its current stage of development, meets the expectations raised in that impressive talk.

Section 2 offers an overview of human computation, the approach behind Duolingo. Then, the author carries out a test-drive of the program – as a reviewer, not as a genuine learner – and presents his observations in regards to
language learning in section 3 and in regards to translating the web in section 4. The concluding remarks in section 5 state the conditions to be met for the Duolingo project to succeed.

2. Human Computation

Luis von Ahn is an associate professor in the Computer Science Department at Carnegie Mellon University and an entrepreneur with successful ventures such as the ESP game and reCAPTCHA behind (Simmons, 2010).

His doctoral dissertation, completed in 2005, coined the term human computation, which will be central to von Ahn’s endeavors including, as we shall see, Duolingo. Human computation aims at combining human and computing power to solve problems neither people nor computers could solve alone. Humans will use computers to do something while computers will use that input to achieve something else. ReCAPTCHA is a good illustration of this. Von Ahn was involved, with Manuel Blum and others, in CAPTCHA, an acronym which stands for Completely Automated Public Turing test to tell Computers and Humans Apart, and accomplishes just that (Ahn, Blum, Hopper & Langford, 2003). With reCAPTCHA he went a step ahead, using CAPTCHA’s typing to help digitise books. ReCAPTCHA does so by presenting two words for typing, one known to the computer, the other not, under the assumption that should the user correctly identify the first, it would have also correctly typed the second; further support to this assumption will be gained once several users agree on the same spelling (Ahn, Maurer, McMillen, Abraham & Blum, 2008). Acquired by Google in 2009, it's being used to identify scanned words that optical character recognition (OCR) software was unable to read.

Linked to human computation is the concept, also popularized by von Ahn, of Games with a Purpose (GWAP), bringing forward game elements to help solving non-game problems. A good example is his ESP game in which two players earn points by assigning the same label to images, with players’ input helping computers’ image recognition (Ahn & Dabbish, 2004). The ESP game was also acquired by Google.

Duolingo, for which von Ahn gives also credit in the TED talk to his graduate student Severin Hacker, is another good example of human computation plus GWAP. Learners gain mastery of a foreign language through translation while their learning exercises are being used to translate the web. Instant feedback is given after each task and “skill points” gained on its successful completion.

Duolingo’s design is pleasant, user friendly, neutral enough to be acceptable no matter which culture or which age. Login takes only and email address (or a Facebook or a Twitter account if you want to carry your community along) and a password. Peer to peer collaboration is encouraged, with Duolingo working at its best when friends challenge each other.

It is still work in progress, with major developments cursorily announced in its official blog – the latest one, an application to use the program on the iPhone (Duolingo, 2012). An Upload Centre was established for those wishing to provide documents for translation. Free of charge for the moment, the plan is to charge for this, if at a very reasonable rate, once the quality of the output can be guaranteed. That’s its business model.

A quick search in a browser will find dozens of self-directed, web-based language learning platforms. Duolingo is free, but so are some others. What sets Duolingo apart is its methodology, based on translation, and its goal, explicitly stated in its login page, of involving the learners in translating the web.

3. Learning a Language

People really can learn a language with it [with Duolingo, by translating sentences]. And they learn it about as well as [with] the leading language learning software.

This is the first bold statement von Ahn made in his TED talk and to which we will pay attention. Translation has a bad reputation in foreign learning and teaching. It is associated with the Grammar Translation Method that for centuries guided the discipline (Larsen-Freeman, 2000: 11-22). This practice disappeared as it was gradually replaced by the communicative methodologies of the 1960s. There are, however, a growing number of voices which argue for its reintroduction. Some consider it as the fifth macro-skill to complement the other four (speaking and listening, reading and writing), which all educated bilinguals, not just translators, should master (Campbell, 2002). In any case, it is a learning method that is proposed for advanced learners (Kaye, 2009) while Duolingo advocates it also for beginners.

Does it work? This article responds on the basis of the author’s test-drive of the program and the reading of some of the comments of genuine learners in its Questions (now renamed Community) tab. The study was conducted first in late October 2012, then in late November, over a few hours on different days, and took the author up to level 7 on the Duolingo’s “skill tree” only.

Translation seems to work, and surprisingly well. Translating is assisted with “hints” when the learner hovers
with the mouse over a word. It gives the learner control over the learning process, as dealing with words only is easier than dealing with the whole communicative situation in which many other factors may apply. Gamification works: the learner feels a sense of achievement when getting the points and challenged when not.

The focus is on reading comprehension. In its Home tab, the learner progresses along a “skill tree” through exercises that mostly involve translation from (and dictation into) the language being learned. This learning is then consolidated on the Translations tab, working always with the language being learned as the source. Some attention is also paid to listening and writing, and to vocabulary and grammatical structures, but none (yet) to oral production.

Most striking is the program’s ability to provide instant detailed feedback at all times, a very effective tool to sustain interest. Feedback on self-directed online learning rarely goes beyond multiple choice. Duolingo on the other hand tells users whether the translation they did or the dictation they copied was correct, and indicates other correct versions if appropriate and the nature of their mistakes if made.

This works quite well in the first lessons, but the more the learner advances, the harder it is for the program to control all variables and, thus, the higher the risk of providing the wrong feedback. The more learners advance, the higher also the chances of them realising they have been provided with the wrong feedback. The program deals with this by prompting “Still think you are correct? Let us know”, but that’s not enough to avoid frustration and annoyance.

The linguistic designers don’t seem to have paid much attention to corpus linguistics and the convenience of presenting words, grammatical structures and fixed expressions based on frequency (Bennett, 2010). Words like alberca, for instance, come well before many other more common and useful. The sentences learners are required to translate or transcribe for practice are often stilted. Readers with some knowledge of Spanish will realize the examples copied below, while grammatically correct, do not ring “authentic”:

Por favor escribe tu libro.
Los gatos beben desde leche hasta cerveza.
Él me escribe un libro.
Mi amigo se fue de mi casa.
Ella lleva comida al restaurante.
Tengo dos semanas de no nadar.

This despite von Ahn’s emphasis on the convenience of actually learning with real content, and setting Duolingo as an example: “As opposed to learning with made-up sentences, [with Duolingo] people are learning with real content, which is inherently interesting,” promises in his TED talk. This certainly happens, however, on the Translations tab. As soon as some content/skill has been drilled, learners are encouraged to proceed to this Translations tab on which they are indeed exposed to the real web content they are meant to help translating.

In his TED talk, von Ahn shows his uncanny ability to make the difficult seem easy:

So the way this works is whenever you’re a just a beginner, we give you very, very simple sentences. There’s, of course, a lot of very simple sentences on the Web. We give you very, very simple sentences along with what each word means. And as you translate them, and as you see how other people translate them, you start learning the language. And as you get more and more advanced, we give you more and more complex sentences to translate.

Some of the texts the author encountered in his test-drive in the Translations tab were excellent choices, well related to the vocabulary and grammatical structures previously practiced. There were, for example, lengthy sentences for which literal word-for-word translation works well, giving the learner, able to deal with unfamiliar vocabulary by hovering with the mouse over the relevant word, a sense of achievement on realising they had successfully completed the task. In some other cases, however, passages were chosen well above the skills learners could reasonably have been expected to achieve.

How to gauge the difficulty of texts for translation is a moot problem (see Hale & Campbell, 2002) and it would indeed be of interest to know how the Duolingo team goes about solving it. Or, to put it in another way, how the team would sort the documents received for translating (or is it the sentences within these documents?), in von Ahn’s words, into the very simple and the more complex.

Once the brief test-drive completed and to gain a better understanding on how Duolingo works, the author browsed through some of the learners comments and insights in the Questions (now, Community) tab.
In most cases, comments show a sense of purpose, users aware they were not just learning a language, but taking part in an exciting experiment, and enjoying it. Learners comprise a vibrant community, volunteering often well articulated insights to developers, who on their part show a good disposition to listen. From these comments we can also infer that gamification works: “I came here to learn Spanish, but I’m staying to gain the points,” writes one. There is mostly enthusiasm, but also the occasional disappointment – “Correct answers marked wrong. Adiós.”

The level of satisfaction seems higher for those working on the first levels, and diminishes as they advance. The same will happen to most language learning packages. Language can be better controlled and feedback can be offered with more precision in the first lessons, as pointed out. Furthermore, beginner learners will notice the difference between having zero knowledge and some knowledge of the language being learned, but their level of excitement will plateau once they don’t feel they advance at the same speed.

The translation approach seems to work, with no one found complaining about it. However, while they are happy with translation, some will feel that not enough material is provided to fully master the skills presented. There are requests, from the developers or from fellow learners, for level-suitable grammar links, podcasts and songs with Spanish lyrics, and movies and TV programs in Spanish with English subtitles.

4. Translating the Web

We combine the translations of multiple beginners to get the quality of a single professional translator.

This is the second bold statement made by von Ahn in his TED talk. Once learners make some progress, they are asked to participate with real texts in real translation projects. Clients paying for these learner’s translations will allow developers to offer Duolingo for free and without ads.

It looks like a typical case of crowdsourcing in translation (European Commission, 2012: 33-34). The translation task is divided in small units (for Duolingo, the sentence) and offered to a pool of volunteers, the language learners, who will take them at their own time and only if they feel confident they can do a good work. They will be well motivated to participate, for the sake of mastering the language, and for the points they will earn and the reputation they will gain amongst their peers. It ticks all the boxes in Howe’s (2008) definition of crowdsourcing. Quality assurance can also be crowdsourced, with fellow learners voting on other translations and even editing them, thus ticking all the boxes in Jimenez-Crespo’s (2009) definition of crowdsourced QA as well.

Duolingo applies here the same strategy Facebook used to localise its site (Mesipuu, 2010: 34-40). It’s not the same situation, however: the Facebook crowd was bilingual already and, already keen users of the site, participants could be considered actual experts on the subject matter they had to translate. The Duolingo crowd on the other hand is comprised of learners, thus functional monolinguals, and not expected to have any particular expertise – although the texts presented to them will be, one would expect, non-specialised, and users do have a choice as for the subject matter they wish to tackle.

The author reviewed Duolingo’s Translations feature at two points, first, end October, and then, end November 2012. The source, as said above, is presented sentence by sentence, with a “View original document” link always at hand. Learner’s translations, in October, were rated in percentages against “current best translation” and learners could translate whole texts no matter how many times a particular sentence had already been translated. This is no longer the case. By end November, depending on the progress made on a particular document, users would receive different prompts. If the document had been just recently posted, they would be asked to “Help us to translate this sentence” with a “computer translation” given as a reference for feedback. If a sentence had already been translated by a few, new learners would be directed to “Rate these translations”. Once enough people had translated (and rated) it and the sentence been given a “100% complete” tag, translating it again was no longer possible; the preferred translation presented still allowed to “Suggest edit” or “Rate translations”, though.

In his TED talk, von Ahn promised a level of quality much higher than that of machine translation. Curiously enough for a computer scientist, he marvelled little about recent advances in translation automation, now capable of producing for some language pairs not yet perfect, nor elegant, but in many cases usable, good enough, fit for purpose translations (Ostler, 2010). He focuses on what machine translation cannot do yet, and promises from Duolingo quality which will be almost undistinguishable from that of professional translators. The example he offers is transcribed in Table 1:
Table 1. von Ahn’s example in his TED talk

<table>
<thead>
<tr>
<th>Input</th>
<th>Reference (20c/word)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falls Pakistans Geschichte ein indicator ist, so könnte Musharrafs Entscheidung, das Kriegsrecht zu verhängen, jener sprechwörtliche Tropfen sein, der das Fass zum Überlaufen bringt.</td>
<td>If Pakistan’s history is any indicator, Musharraf’s decision to impose martial law may prove to be the proverbial straw that breaks the camel’s back.</td>
</tr>
<tr>
<td>Duolingo</td>
<td>If Pakistan’s history is an indicator, Musharraf’s decision to impose martial law could be the proverbial straw that breaks the camel’s back.</td>
</tr>
</tbody>
</table>

For this example, the quality of Duolingo’s final version can certainly be considered as undistinguishable from professional translation.

In October 2012 the author practised with several texts in the Translations tab, and details for a particular sentence are reproduced in Table 2. The first author’s attempt at translating the Input sentence got a disappointing feedback. Trying to understand how this instant, automated feedback worked, he entered the source into Google Translate and pasted the machine’s version in Duolingo, this second time getting “94% agreement with correct solutions from others”. For this particular example, Duolingo’s “current best translation” can certainly be considered as undistinguishable from … machine translation!

Table 2. Author’s example

<table>
<thead>
<tr>
<th>Input</th>
<th>Human translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconozco que la calculadora iPod canta un poco pero el podómetro Shuffle está muy conseguido (aunque sea un poco más grande)</td>
<td>I admit the iPod calculator sucks a bit, but the Shuffle pedometer turned out quite well (even though it’s a little bigger).</td>
</tr>
<tr>
<td>Google Translate</td>
<td>I recognize that the iPod calculator sings a little but the pedometer is very accomplished Shuffle (even slightly larger).</td>
</tr>
<tr>
<td>Duolingo (“current best translation”)</td>
<td>I recognize that the iPod calculator sings a little but the pedometer Shuffle is very achieved (even if it is a bit more large).</td>
</tr>
</tbody>
</table>

Duolingo seems to have machine translation as the reference on which it bases its automatic feedback. Then they hope learners, by voting and editing, will end up getting the final version right. That’s what must have happened in the German example in Table 1 – which must have had the professional version entered as a learner’s input.

At the time of writing (end November 2012) and for the particular example in Table 2, Duolingo’s “Show final translation” was still that shown, with no progress made towards improving its rendition despite the weeks passing and the involvement of 30 people in its translating, editing and rating.

Even if for the Table 2’s particular example Duolingo doesn’t seem to have (yet) worked well, the system looks clever enough, with many other “final translations” indeed remarkably good taking into account authors’ assumed knowledge. Although we can all access all prior translations of a particular sentence for the purpose of rating and editing, it is not clear from the front-end how the decision on final translations is reached at. Provided some have translated the whole sentence well (or even just chunks of it), it should be theoretically possible to manage, through learners’ editing and rating, for the best possible solution for that sentence to emerge. The right seeds need to have been planted there for each sentence, though, and that cannot be guaranteed. Then, even if the best available option could emerge at the sentence level, there doesn’t appear to be any mechanism yet to ensure coherence across sentences.

5. Will It Work?

It is reasonable to expect Duolingo could eventually work, but only for texts for which accurate and elegant
translation was not critical, texts any educated bilingual without specific translation training could translate. One could expect from Duolingo an output at a level of quality similar to that achieved by postediting machine translation. Duolingo seems to encourage a literal (almost word-by-word) approach to language transfer, the same approach machine translation adopts. The advantage over machine translation, however, is that learners can add common sense and knowledge of the world. We find ourselves again in typical human computation territory, with human intelligence filling those gaps computer processing can’t yet.

To make this possible, however, Duolingo will need to retain massive numbers of advanced language learners, able to provide for each sentence (or for each chunk) at least one good quality version other users will pick up while editing and rating. This is not easy. The rate of attrition for language learning – the index between those who start learning and those who exit with a reasonably good command of the language learnt – very high on face-to-face tuition (Bardovi-Harlig & Stringer, 2010), must be much higher in self-directed online learning. Without massive numbers of advanced students, Duolingo won’t be able to make charging for translation a sound commercial proposition.

To retain those massive numbers of advanced learners, Duolingo will need first to finetune its current instant feedback algorithms. Even within the controlled environment of a learning package, to allow for all possible correct variants of a given source while pointing out the non correct ones is not an easy task, particularly once the Language 101 stage is left. That alone, however, won’t be enough. It will need also to capture the whole online language learning market, to be for online language learning what Google is for search or Facebook for social networking. That’s not easy either. There is a fierce competition out there. A quick search soon found over forty hits for self-directed language learning platforms and the list is not exhaustive. (Note 1)

Duolingo lacks many of the bells and whistles available in state-of-the-art computer assisted language learning (CALL). To have a chance to capture the whole market it will need to add, to what is a very good foundation, audiovisual context as presented for example in BBC Languages, sophisticated uses of advances in speech recognition and synthesis to practice pronunciation as Rosetta Stone does, use of virtual words as in Avatar English, interaction with the learning crowd not only based on peer-to-peer, but also on native-to-learner contact as in Livemocha, and much more.

Duolingo works on the same human computation principle as reCAPTCHA. Technology-wise, taking the sentences from a document in the web, translating them in Duolingo and inserting their target version back into the target document seems less complicated than what reCAPTCHA does: identifying the words OCR can’t read, presenting them to users, then placing them with the right spelling in the document from which they were taken. Translation however is not about words or even about sequences of words, it’s about meaning. And ultimately, this is where Duolingo may fall short.

It’s work in progress, but developers must be under enormous pressure to make the system work before funding dries up. It may succeed, and even become Google’s next acquisition target (after the ESP game and reCAPTCHA) as Siegler (2011) has tongue-in-cheek suggested. A plan B, should it not succeed as a commercial venture, could be to open its source and let the crowd at large extend it and improve it: the start of a rich web-based language learning platform of Wikipedia proportions!

References
Ann Arbor: University of Michigan Press.


**Note**

*Note 1. In alphabetical order: Avatar English; BBC Languages; busuu.com; Conversation Exchange; elanguageSchool.net; en.bab.la; espod.com; fluentfuture.com/friendsabroad.com; Gabble-on.com; HappyTutors.com; Hello! Mandarin; iKnow!; ingolingo.com; Interpals; italki.com; Japancast.net; lancelotsschool.com; Language Buddy; Language Exchange Network; Lingozone; Livemocha; Mango Languages; Mixxer; MyLanguageExchange; myngle.com; Palabea; Soziety.com; spanishdict.com; Spanishpod - Frenchpod – ChinesePod; speakshop.com; talkconnigo.com; talkingbookslibrarian.blogspot.com.com; TT4You; underspeak.com; Vocabulix; wordchamp.com; wordcurious.com; worldwordexchange.com; xLingo; yabla.com.*