Modelling memory with Busuu’s Vocab Trainer.

April 2019: Data Science Festival
Intro/Agenda

Kirsten Campbell-Howes
Head of Education

- 18+ years in the language learning and edtech industry
- Masters in Applied Linguistics from University of Oxford

- Global language learning market
- About busuu
  - History
  - Product
- Data Science at busuu
  - Churn
  - Vocabulary Trainer
  - Roadmap
- Q&A
The global language learning market
Digital language learning to explode in the $54bn market

Worldwide Language Learning Industry

- Total target market: $54bn+
- Online language learning: 8%

Global Digital Language Learning market

- CAGR: 19%
- 2016 to 2021
- English learners worldwide (billions)
  - 2010: 1.0
  - 2020: 2.0
  - Growth: 2x
busuu has >90 million users growing at 40,000 users per day

Our top markets in 2018
We have a freemium business model with attractive Premium benefits

<table>
<thead>
<tr>
<th>Free</th>
<th>Premium - €10/month*</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ 1 language</td>
<td>✓ 12 languages</td>
</tr>
<tr>
<td>✓ Flashcards</td>
<td>✓ Flashcards</td>
</tr>
<tr>
<td>✓ Writing exercises</td>
<td>✓ Writing exercises</td>
</tr>
<tr>
<td></td>
<td>✓ Speaking exercises</td>
</tr>
<tr>
<td></td>
<td>✓ Grammar units and interactive quizzes</td>
</tr>
<tr>
<td></td>
<td>✓ Intelligent Vocabulary Trainer</td>
</tr>
<tr>
<td></td>
<td>✓ Faster correction from native speakers</td>
</tr>
<tr>
<td></td>
<td>✓ Offline mode on mobile devices</td>
</tr>
<tr>
<td></td>
<td>✓ Certificate</td>
</tr>
<tr>
<td></td>
<td>✓ Travel course and Business Course</td>
</tr>
<tr>
<td></td>
<td>✓ Personalized Study Plan</td>
</tr>
</tbody>
</table>
Our history
2008

2019
busuu product and pedagogy
Our vision is to Break down language barriers
Our mission is to provide the closest experience to living in another country through intelligent learning enriched by interactions with native speakers.
busuu combines AI-powered courses with instant feedback from native & advanced speakers

1. Al-powered language courses
   - High quality courses in 12 languages* Created by experts
   - Intelligent learning with AI (Predictive vocab trainer, adaptive study plan, speech recognition etc.)

2. Practice with native speakers
   - 90m strong user community from 190 countries
   - Instant feedback on voice and text exercises

* English, Spanish, German, French, Italian, Portuguese, Mandarin, Russian, Japanese, Arabic, Polish, Turkish
We are one of the most complete language learning apps in the market
Our methodology

busuu is a complete self-study and language practice environment

- Vocabulary
- Grammar
- Pronunciation

- Listening
- Reading

- Writing
- Speaking
- Interaction
Describe cómo era un día típico de clase para ti.

Kirsten Campbell-Howes
2018-09-12 10:19

Cuando era joven, la asignatura que me gustaba más era inglés. Me siempre escuchaba bien en clase y mis profesores creían que era un buen estudiante.

Isabel Esquivias
2018-09-12 10:21

Cuando era joven, la asignatura que más me gustaba era inglés. Me siempre escuchaba bien en clase y mis profesores creían que era un buen estudiante.
We have expanded our products into machine learning, Virtual Reality and bots.
Academic and research partnerships

2018: busuu wins the ‘Comenius Award’, the most prestigious award in Germany for high quality educational products

2018: Study with The Open University shows that 82% of busuu users believe they improved greatly their language knowledge while studying with busuu

2018: busuu wins the ‘EdWard Award’ for high-quality learning product from University College London

2016: Efficacy study with City University of New York suggests that 22 hours of busuu is as effective as an entire college semester in the US
Data Science
Data science at busuu

> **84k users** complete an exercise each day

> **4.5M graded exercises** per day = 3,125 in the last minute!

---

**Dr Thomas Richardson**  
Senior Data Scientist

- 4 years as a Senior Data Scientist specialising in ML
- PhD in astrophysics from King’s College London

**Dr Stephanie Reynolds**  
Junior Data Scientist

- Research experience in Statistics, online retail and Biophysics.
- PhD in mathematics from Imperial College London
Study plan

Take the stress out of learning French
with a personalised Study Plan that fits around your life

- Choose your goals
- Set your schedule
- Generate a Study Plan
- Work in small, achievable steps to a completion date

Start Study Plan

What is your main goal in learning French?

- Feel like a local when I travel
- Develop professionally
- Help with my education
- Learn for fun and culture
- Communicate with friends and family

What level do you want to achieve?

- BEGINNER
  - Order a meal in a restaurant and understand signs
- ELEMENTARY
  - Make small talk with locals and talk about my life
- INTERMEDIATE
  - Find my way around on my own
- UPPER INTERMEDIATE
  - Understand local radio and TV

Which days of the week would you like to learn?

- Monday
- Tuesday
- Wednesday
- Thursday
- Friday
- Saturday
- Sunday

Continue
User engagement

The motivation of students can decrease over time, *how* and *when* can we intervene to help users achieve their language goals?
User engagement: effect of motivation

Some users may be inherently more engaged than others, due to their motivations for learning a language.
Users that have recently completed a study plan tend to be more engaged on busuu. They both

- complete more lessons,
- send more exercises to be corrected by the busuu community.
Modelling vocabulary acquisition

Agenda

- busuu’s vocabulary acquisition strategy
- Key lessons learned
  - Review exercise design principles
    - What does it mean to ‘know’ a word?
  - Modelling vocabulary strength and its decay
    - Which data best predicts when users forget a word?
- Ideas for the future
Learning vocabulary with busuu

- First interaction with vocabulary in the Lesson tab
- **Human-designed** content provides context for each word.
- Vocabulary introduced according to user preference and CEFR levels
Learning vocabulary with busuu

Memorise with flashcards

Tips for context

First tests
Reviewing vocabulary with busuu

- Can see current vocabulary in the review tab
- Assign vocabulary strength of weak, medium or strong
- User can choose to review “weak” words only = personalised learning

Your vocabulary

<table>
<thead>
<tr>
<th>Spanish</th>
<th>English</th>
<th>Word strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me llamo...</td>
<td>My name is...</td>
<td>Medium</td>
</tr>
<tr>
<td>Más o menos.</td>
<td>So-so.</td>
<td>Medium</td>
</tr>
</tbody>
</table>
Measuring strength

We used the **Area under the ROC curve (AUC)** to gauge model performance of scores $P_W$ as a means of ranking future review exercises by difficulty.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuoLingo$^1$ Leitner</td>
<td>0.542</td>
</tr>
<tr>
<td>DuoLingo$^1$ Half-life Regression 2016</td>
<td>0.537</td>
</tr>
<tr>
<td>DuoLingo$^2$ LR 2018</td>
<td>0.774</td>
</tr>
<tr>
<td>busuu LR</td>
<td>0.780</td>
</tr>
<tr>
<td>SANA Labs$^2$ Recurrent nets + GBT Ensemble</td>
<td>0.861</td>
</tr>
</tbody>
</table>

$P_W$: the probability that a user will successfully remember word $W$ when they next practice.

---

Review exercises: Design principles

1. User experience
2. Impact on learning
3. Measures / predicts vocabulary strength
   - A new requirement for an automated ML-based review system

Score exercise types based on these three principles
Review exercises v1.0

User experience

Impact on learning

Measures user progress

Multiple choice exercise
Multiple choice exercise

Select the correct option.

fröhlich

the piano  happy  My name is...

Without intelligent distractors average pass rate of 95%

Masks actual knowledge of a word. Class imbalance problem for supervised learning of vocab strength.
Review exercises v1.0

User experience

Impact on learning

Measures user progress

Multiple choice exercise

fröhlich

the piano  happy  My name is...
Review exercises v2.0

L1 to L2 translation exercise

Dictation exercise
Review exercises v2.0

User experience
Impact on learning
Measures user progress

L1 to L2 translation exercise
Dictation exercise
Translation exercise:
Mistakes by type

Why is the first-time pass rate so much lower for some words? Mistakes can be:

- Missing accents
- Alternative answers
- Minor typographical errors
- Punctuation

Just 47% of errors were major (as indicated by large ratio of Levenshtein distance to total character length)
Review exercises v2.0

User experience

Impact on learning

Measures user progress

What does it mean to *know* a word??
Flashcard experiment - what do users say they know?

User experience

Impact on learning

Measures user progress

Self-marked flashcard exercise
Flashcard experiment

- For the majority of words more than 85% of users self mark that they know a word.

- The majority of words had significantly higher pass rates for flashcards than translation. The Pearson correlation coefficient in mean pass rates for flashcard and translation exercises was $p=0.39$. 
Flashcard experiment

More translation practices

More flashcard practices
Flashcard experiment

Self-marked flashcard exercise

User experience

Impact on learning

Measures user progress
Looking ahead: stages of knowing a word

<table>
<thead>
<tr>
<th>Stage</th>
<th>Evidence</th>
<th>Recommended exercise</th>
<th>busuu strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner</td>
<td>Fails test on receptive exercises. Self-marks as unknown on flashcards. Major failures on interface to course translation exercises.</td>
<td>Flashcards / course to interface MCQ with intelligent distractors</td>
<td>Weak</td>
</tr>
<tr>
<td>Recognises word</td>
<td>Marks flashcards as known. Passes course to interface MCQ with intelligent distractors. Minor typos on interface to course translation exercise.</td>
<td>Interface to course language translation exercise so that they can get the exact spelling.</td>
<td>Ok</td>
</tr>
<tr>
<td>Mastered word</td>
<td>Perfectly recalls word on interface to course translation exercise.</td>
<td>None. Wait until the word decays again over time and test with interface to course language again later.</td>
<td>Strong</td>
</tr>
</tbody>
</table>
Research indicates that practice sessions should be spaced at longer and longer intervals as a user builds familiarity with a word.
First system used by busuu is similar to the **Leitner system**

- Decay rate takes levels $L=1,2,3,4,5,\ldots$
- $L$ is a half-life in days. If your last practice was $t$ days ago your strength is $s = 2^{-L/t}$
- If you are tested on a word and remember it correctly your level increments $L+=1$
- If you get it wrong your level $L-=1$

**Strengths**
- Personalised learning: Words you fail decay faster.
- Simple and easy to understand. Users get consistent behaviour.

**Weaknesses**
- All words treated the same.
- Cold start. Require lots of interactions before confident that user knows a word.
- Competitors have demonstrated with A/B test that ML model outperforms Leitner on repeat sessions (+12%).
Measuring strength

We used the **Area under the ROC curve (AUC)** to gauge model performance of scores $P_W$ as a means of ranking future review exercises by difficulty.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuoLingo$^1$ Leitner</td>
<td>0.542</td>
</tr>
<tr>
<td>DuoLingo$^1$ Half-life Regression 2016</td>
<td>0.537</td>
</tr>
<tr>
<td>DuoLingo$^2$ LR 2018</td>
<td>0.774</td>
</tr>
<tr>
<td>busuu LR</td>
<td>0.780</td>
</tr>
<tr>
<td>SANA Labs$^2$ Recurrent nets + GBT Ensemble</td>
<td>0.861</td>
</tr>
</tbody>
</table>

$P_W$: the probability that a user will successfully remember word $W$ when they next practice.

Cold start: Helps problem from 1.0. Historical data to learn that “hi” is easier than “to lie down”

Spaced Repetition: Ensures that strength increases with successive passes

Words failed often decay faster

\[
p_w = \sigma \left( \theta_w M_w + \theta_p N_p + \theta_f N_f + \theta_0 \right)
\]

- \( M_w \): The mean pass rate on translation exercises for word \( w \) across all users.
- \( N_p \): The number of times that the user has passed any graded exercise on word \( w \)
- \( N_f \): The fail rate for the user on graded exercises with the word \( w \).
## Difficulty by language pair

<table>
<thead>
<tr>
<th>Language</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Literature</td>
</tr>
<tr>
<td>Spanish</td>
<td>Literatura</td>
</tr>
<tr>
<td>Polish</td>
<td>Literatura</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Literatura</td>
</tr>
<tr>
<td>Russian</td>
<td>Литература</td>
</tr>
<tr>
<td>French</td>
<td>Littérature</td>
</tr>
<tr>
<td>German</td>
<td>Literatur</td>
</tr>
</tbody>
</table>

### Beta distribution

\[
f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1 - x)^{\beta-1}
\]

*busuu 2019*
Spaced repetition based on performance and difficulty
1. Use $P_w$ to set the half-life $h_w$ of word strength decay rate.

1. A user’s word strength $S_w$ will then decay over time so that after $t$ days since last practice

$$S_w = 2^{-h_w/t}$$
DS Roadmap

Projects in the pipeline

- Grammar review
- Fluency metric
- Mistakes classification and remediation
- Automated corrections (with human in the loop)
- Phoneme review

Your Grammar

<table>
<thead>
<tr>
<th>Weak topics</th>
<th>Medium topics</th>
<th>Strong topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Favourite topics

- Pronouns
- Verbs
- Masculine and feminine
- Singular and plural

Verbs

- Ser, estar, the different tenses, regular and irregular verbs, etc
- Ser and estar
  - Reference
  - Practice
- Presente simple de indicativo
  - Reference
  - Practice
- Presente simple regular verbs
  - Reference
  - Practice
Q&A

kirsten@busuu.com

Thank You