※ To view images in stereo using anaglyph glasses, please refer to the colour version available online at the Tsukuba Repository or to a colour printout thereof.

RUDE Markus HALL William

Vision and Language - the Advantages of Binocular Vision for Foreign Language Acquisition

Abstract

The evolution of our visual system outdates the existence of written language. According to Zeki (2003), our ability to extract information from written language is underdeveloped compared to that from visual information. This paper explores new ways of utilizing binocular vision for language learning. Three-dimensional (3D) text representations allow for the grouping of parts of text, providing a new method of controlling text redundancy. Binocular Rivalry (BR), the perceptual alternation between different images presented to both eyes simultaneously, is explored as a way of enriching bilingual texts and as an alternative to vocabulary memorisation tools such as flip-cards. 3D and BR could be beneficial for foreign language teaching texts with enhanced information (glosses, pronunciation cues etc.) or reduced redundancy (cloze tests, gap-filling tasks etc.). Visual examples of suggested, or actual, language-teaching applications are shown using anaglyph images and free-viewing stereograms, and preliminary experimental results are reported. Limitations and possible solutions for visually challenged students are discussed.

[Keywords] *Vision, foreign language learning, 3D, binocular vision, binocular rivalry (BR), redundancy of texts*

Introduction

Neurologist Semir Zeki claims (2003) that our ability to discern meaning from language is underdeveloped when compared to our ability to discern meaning from visual information. He argues that the superiority of the visual system, having evolved over many millions of years longer than the linguistic system, enables a much more efficient method of recognition. In this paper we propose that the way in which we use both eyes to see offers us several interesting and unexplored possibilities in terms of foreign language education.

Vision plays a key role in foreign language learning. After all, one of the first steps in the complex cognitive process of reading is the act of seeing itself. To derive meaning, we first need to decode the visual symbols presented to our eyes. Through practice we become skilled at reading, and the process becomes automatic. However, it is worth remembering that both our ability to read and to see are learned in childhood. Each process developed out of a complex combination of cognitive, perceptual, and social learning, which were either taught by structured systems or acquired automatically (Rayner, et al., 2001). In essence, seeing and reading are both learned processes of constructing meaning from visual information.

As is the nature of evolution, incremental developments are unnoticeable in a single lifetime. Reading, on the other hand, is a relatively recent development, and inventions such as the digitalization of text, and the ubiquity of personal handheld devices, have drastically altered the way meaning can be acquired from text. We believe that alternative methods of reading text, when applied to established structures of the visual system, may reveal advantages for language learning.

We will explore two new ways by which the characteristics of binocular vision can be used to help foreign language learning. The first is the use of three-dimensional (3D) representation to exploit our tendency of *grouping* (Wolfe, Kluender, & Levi, 2020) visual information. Second is the use of *binocular rivalry* (BR), which is the experience of perceptual alternation between different images presented to both eyes simultaneously, to hide and reveal different language texts. Possible classroom applications and limitations will be discussed.

Seeing with two eyes

Due to the horizontal separation of the left and right eye, images projected onto the retinae are slightly different. Providing the images are similar enough, they can be fused together to perceive depth by a process known as *stereopsis* (Howard & Rogers, 1995). If the disparity is too large, then they cannot be fused, and one image may be temporarily suppressed. While we are generally unaware of these background processes in our daily lives, they have huge implications in terms of how we experience our environment and are of great interest to researchers curious about the nature of our perceptual experience.

Humans have two eyes yet *see* only one image. Often taken for granted, this ability to create a single 3D experience from two-dimensional images projected onto the back of each eye is one of the most remarkable aspects of human vision. Being able to see in stereo offers several profound evolutionary advantages, including refined depth perception and camouflage breaking (Wilson & Julesz, 1971).

Primitive explanations of how vision works date back to the ancient Greeks, yet it was not until relatively recently that we began to truly understand the mechanism by which we see in 3D, largely due to the 19th Century invention of the stereoscope (see Fig. 1). Charles Wheatstone's device, used for displaying and experiencing three-dimensional images, helped elucidate the mechanism of 3D vision by making *seeing* itself the subject of observation.

Advancements in technology over the last 150 years, and the digitalization of photography and film, have greatly changed the methods used for reproducing 3D vision. Furthermore, contrary to the aims of early research, modern 3D technology endeavours to mask the process by which the image is acquired.

Contemporary devices such as 3D glasses and virtual reality headsets are designed to create the most convincing experience possible by drawing our attention away from *how* we see and towards *what* we see.

Although our ability to demonstrate 3D vison is remarkably recent, and the materials and methods used to simulate the experiences have changed dramatically over the last century, the physiology of 3D vision itself has evolved at a much slower pace since our early human ancestors (Fox, 1978). The exponential advancements in technology, and ever-deepening understanding of how vision works, suggest increasing potential for technological intersections with our sensory systems.

Note on stereograms in this paper

Most stereograms throughout this paper are overlaid red and cyan images and are viewed using anaglyph glasses, with the exception of a few free-viewing stereograms. The red and cyan lenses of the glasses allow different images to pass. Subsequently, each eye is shown a different image. If the images are within a certain threshold of similarity, the brain fuses them together and we perceive three-dimensional space (see Fig. 2). Anaglyph stereograms were used for two reasons. Firstly, the glasses needed are cheap and easily accessed by readers. Secondly, the comparatively simple nature of the anaglyph 3D set-up makes it easy to understand what is happening. The same principles of vision apply to all ways of reproducing 3D, such as virtual reality (VR) headsets, which may be more suitable for educational purposes due to the minimal distraction of the apparatus and focus on the experience itself.

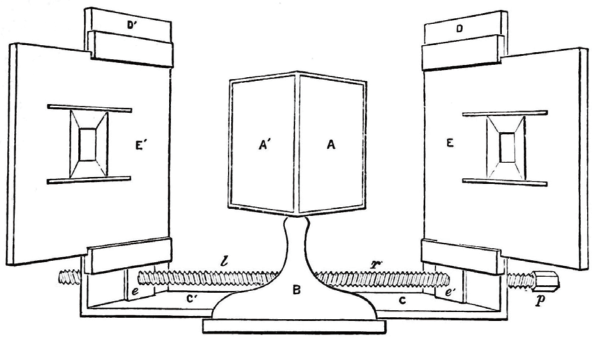

Figure 1: ***Wheatstone Stereoscope, 1838. Viewers look through mirrored viewing device (A) at stereograms positioned (E). Left and right eyes are shown different images simultaneously which the brain automatically attempts to fuse together. The original design, intended for vision research, aimed to unravel the complexities of vision.***

Figure 2: ***3D image demonstrated by overlaid photographs of a corridor taken 6cm apart horizontally. When viewed with anaglyph glasses, the left and right eye see what the viewer would were they to stand in the corridor. As in real life, the brain fuses the slightly different images together to create a 3D experience.***

Reduced and increased redundancy

How does binocular vision relate to language learning? It is worthwhile to consider one typical visual "object" in language teaching – a visualized text – and to discuss the amount of information it contains.

Language teaching uses texts with *enhanced information* (word translations, pronunciation cues etc.) and with *reduced information,* also termed *reduced redundancy* (cloze tests, gap-filling tasks etc.). 3D and BR can help create such information enhanced or redundancy reduced texts.

Texts for native speakers (NS) have a redundancy of about 50% (on character level, for English, Shannon & Weaver, 1949) or even more (if word or text level redundancy is also considered, and there is similar redundancy for other languages). This means that every second new character can be predicted from previous information (up to a certain character) and is thus seemingly superfluous. This seems like a waste of information being exchanged; however, it is not: it allows reliable communication even in the presence of background noise.

Texts for language learners are often equipped with *glosses* (word translations, but also paraphrases, pronunciation cues, etc.), which means that they have *increased redundancy* compared to the pure text; glosses are superfluous for a native speaker (thus redundant), but essential for understanding for a language learner; for him or her, the texts are information enhanced (not redundant, since unfamiliar words are not yet in the mental lexicon). On the other hand, for tests or repetition exercises, gap-filling tasks are often used. In that case, we have the situation of a text being stripped of a part of its information, or – in other words – with *reduced redundancy.* In fact, if there was no redundancy, and every character and word contributed absolutely new information (non-deducible from the rest of the word or text), then gap-filling tasks would be unsolvable, even for NS.

Using 3D vision for glossing

When learning to read in a foreign language, adequate learning materials are essential. Glosses are a big help in this task. Salem (2006) investigated the influence of various types of glosses (Spanish native speakers learning English) and concluded that "[t]he outcomes of this study support the need to include glosses in foreign language reading texts, since it has a positive influence on reading comprehension and word retention." The renowned vocabulary expert Paul Nation had suggested word explanations far from their appearance in a text, e.g. at the end of a book, in order to stimulate guessing (Nation 1990, 134). However, readers prefer them near, at least on the same page. Nation (2008, 62) seems to have compromised and writes: "Learners prefer the glossary to occur next to or near the line of text containing the glossed word."

The closest possible gloss location is just above or below the target word. This style of glossing, in-between the lines, is called *interlinear glossing*. In sciences, the number of explanatory glosses (additional lines) might be quite high, in language education, there are mostly just one or two. In Hiragana Times (2020), a Japanese-English bilingual magazine, several types of glosses are used with a varying number of interlinear glosses; some texts have three additional lines, e.g. for (1) reading (Furigana), (2) reading (Romaji), and (3) word or chunk translations. The English translation (3) is sometimes in red such that readers can hide them by using a red colour filter. However, the former two glosses (1) and (2) are rather close to the target text and always visible, thus more separation could be desirable to avoid distraction which might be difficult to realize due to space limitations.

Now, having the third dimension (depth) at the disposal of the reader offers a combination of proximity and distance: binocular vision acts like a natural filter without the need of a manual operation (e.g. applying a red filter, clicking the mouse button, turning a flash card) to access or hide the gloss: access is granted simply by vergence (turning the eyes away from or towards each other, i.e. divergence or convergence; see Fig. 3 while using anaglyph glasses).

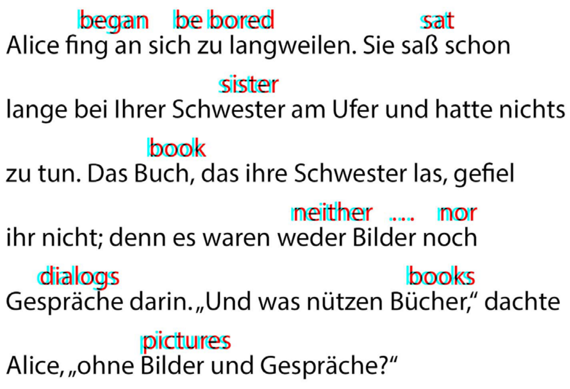

Figure 3: ***3D glossing example with an anaglyph stereogram: main text (German) on the screen layer (layer I), translations one layer behind (layer II). The German target text is – slightly adapted – from the German version (Carroll, 2007) of Alice's Adventures in Wonderland.***

Making the glosses deeper forces the reader to focus his/her attention on a plane of a very different depth and has two effects. First, the target text is *grouped* more strongly relative to the now more seperate group of glosses, the so-called grouping principle of proximity, here in depth, applies. (Others: the grouping principle of similarity, symmetry, parallelism, etc., Wagemanns et al., 2012) Eye movements from line to line are therefore easier since the eyes are not as much distracted by the interline gloss as if it were on the same or nearly the same depth layer. But second – more important – the glosses are perceptually further away and thus “hidden”: the reader needs to intentionally diverge and might perceive them only after consciously gazing at them, there is a smaller chance of reading the glosses unintentionally; this in turn might stimulate guessing and thus satisfies both of Nation’s conflicting demands, having glosses distant enough for a guessing need, but close enough for easy access.

By varying the depth, this second effect (hiding the glosses, as long as the target text is read) is intensified in Figure 4.

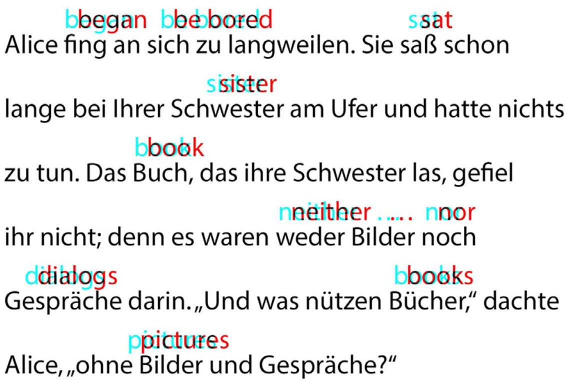

Figure 4: ***3D glossing example as in Fig. 3 (anaglyph image), here with more depth: thus, more divergence is needed to access the glosses. The higher effort of access goes along with a clearer separation of depth layers, and thus likely with less distraction by the glosses while reading the target text.***

The use of depth as a text layout option makes it possible to include several lines of glossing on different depth layers. Glosses of one semantic category can be perceived as independent visual units, clearly separated not only from the target text, but also from each other – only one depth layer with its glosses is clearly visible at a glance. Interline spacing (in 2D, in the vertical) can be reduced, since there is additional “virtual interline spacing” (in 3D, in depth) which serves the same purpose of segregating the text elements from each other. As an example, the English translation of the whole text is added on an additional depth layer in Fig. 5, further back than the word glosses.

Even for readers without anaglyph glasses, looking at the text gives an idea of the 3D perception when verging on the German text: it is clearly readable whereas the keyword glosses (close behind) look somewhat smeared. To each red character, there is an overlapping bluish “twin” making it hard to read. Almost unreadable, however, is the English text in the smallest font. (This applies to a hard copy. The effect might be less noticable on a screen.)

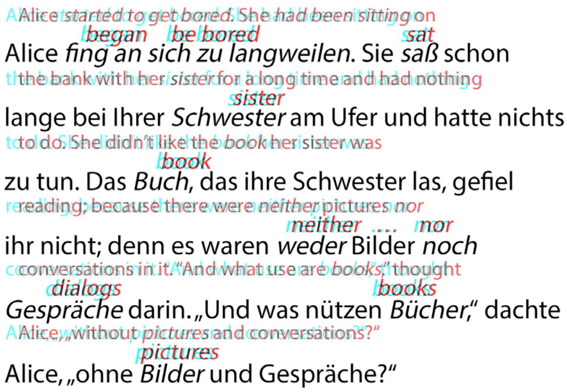

Figure 5: ***A 3D glossing example with three depth layers: target text again on screen layer (depth layer I), translation of keywords close behind (depth layer II), and complete English translation further behind the main German text (depth layer III).***

Discussion and uses in class

Glosses as the ones above have not been used yet in the classroom; however, text examples with reduced redundancy, gap-filling puzzles, have. One such prototypical application involves both, the same text with reduced and increased redundancy (enhanced information). The difference of text redundancy is made by switching from monoscopic to stereoscopic view: the switch allows to reconstruct the original text (normal redundancy); moreover it also makes sentence stresses visible (increased redundancy for NS, enhanced information for language learners. See subsection *Actual classroom application II*.)

Another classroom use is a text with phonetic cues (increased redundancy) in a writing style called Prosodic Writing (PW), showing intonation and sentence stress (Rude, 2016). The versions of PW used in class were only pseudo-3D forms of text, but one student contributed an experimental 3D version of PW, a computer-generated free-viewing stereogram. (See Appendix.)

Binocular Rivalry

Through stereopsis, information in both eyes is constructed into a coherent 3D impression. The phenomenon of BR, however, occurs when the images presented to each eye are too different to be fused. The brain struggles to create a meaningful interpretation of the two views and one image will temporarily dominate our perception, while the other is suppressed. So long as the strengths of the images remain the same, our perception will switch back and forth between the two in a bistable perceptual experience (Alais & Blake, 2005) (Fig. 6).

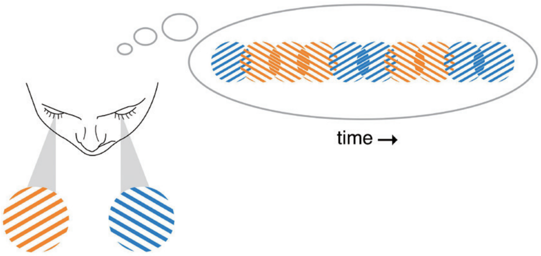

Figure 6: ***BR stimulus and experience. Different images viewed simultaneously with each eye will battle for dominance of conscious awareness.***

BR is most commonly demonstrated using different overlaid images (see Fig. 7). When viewed using anaglyph glasses, each eye is shown a different image. Despite *seeing* both images, in the sense that they are each projected simultaneously onto the retinae, only one tends to dominate awareness at any one time, while the other is suppressed. We commonly experience a sequence of perceptual switches every few seconds as the brain struggles to resolve the conflict. One fascinating aspect of BR is that our subjective perception seems to flip back-and-forth between the two images, despite the images remaining unchanged. By alternately closing our left and right eye, while using anaglyph glasses, we can easily prove that we do in fact *see* both images.

Figure 7: ***BR demonstrated using two different images. The red (R) colour channel is used for one image, and the green (G) and blue (B) are used for the other. When viewed with anaglyph glasses, different images are shown to the left and right eye. The brain struggles to create a coherent experience and switches back-and-forth between the two.***

BR and language

If we consider text as visual symbols from which the brain discerns meaning, then it is a reasonable assumption that a similar BR experience can be produced if different texts were shown to each eye. By overlaying English and Japanese text (see Fig. 8) it is possible to produce such a bistable experience in which different language texts battle for perceptual dominance. The reader may notice that the rivalry between the text is not as clearly defined as in Figure 7. While the BR experience of Figure 7 is mostly binary, a more complex battle occurs in Figure 8. Although one single text does generally tend to dominate, we rarely see one text exclusively. The experience is more fragmented and ambiguous. The suppressed text, although less noticeable, maintains a presence in our mind.

One possible explanation is that the images in Figure 7 are easier to distinguish from one another than the text in Figure 8. As our eyes trace the form of a cup and book on a table-top, the holistic integrity of the images and our prior knowledge of familiar shapes may cause us to expect the remaining image: a concept known as "priming" (Hoffman, 2000). The text in Figure 8, on the other hand, is harder to distinguish. As we recognize the contours of the Japanese text, for instance, we may mistakenly group together a section of English. Although we may assume that Japanese and English characters are visually different enough to immediately group them separately, perhaps this assumption comes from our prior knowledge of the languages themselves as fundamentally different, and, in actual fact, the forms of the characters are more similar than we imagine. Certainly they are visually closer than the rivaling images in Figure 7. Before discussing how this issue may be resolved, it is useful to explore the nature of the BR mechanism in more detail.

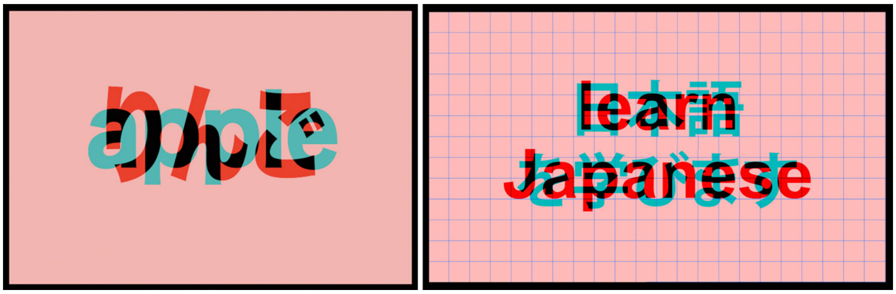

Figure 8: ***BR experience of overlaid single and multiple-word Japanese and English text. By the same method of editing the colour channels and viewing the images with anaglyph glasses, Japanese and English text can be shown to the left and right eye simultaneously.***

Which image is more interesting?

"Stimulus strength" plays a key role in determining which image we are consciously aware of in a BR experience (Levelt, 1968). Simply put, a visually interesting image is more likely to dominate conscious awareness and a less interesting image is more likely to be suppressed. Both high-level, cognitive driven factors, such as meaningful content, and low-level, stimulus driven factors have been identified as influencing BR. The key low-level factors identified are brightness, focus, and movement (Levelt, 1968) (see Fig. 9).

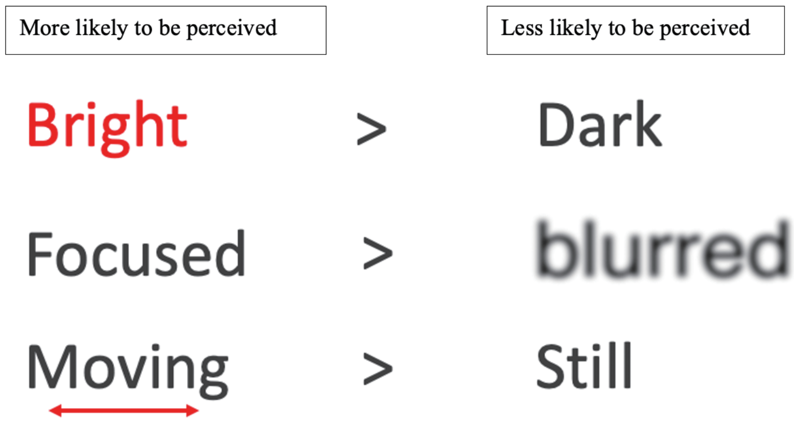

Figure 9: ***Stimulus strength affects which image dominates conscious awareness in BR.***

One study (Hall, 2020) tested the possibility of controlling the dominant language in a BR experience by adjusting the stimulus strength of Japanese and English texts shown simultaneously to each eye. Participants (n=26) from an undergraduate university in Japan viewed 20 GIF images of overlayed English and Japanese text displayed in either red or cyan. When viewed through anaglyph glasses, each text would appear at the same retinal position in either eye, creating a BR experience. In order to make one language text dominate over the other, a lateral motion effect was applied to one of the texts in each image (<0.5% of total text area per 100 ms) (see Fig. 10). The images were shown for 10 s each and participants reported which language they perceived (*Japanese only*, *mostly Japanese*, *English and Japanese equally*, *mostly English*, or *English only*). This preliminary study found a positive correlation between movement and dominance of conscious awareness. When the motion effect was applied only to Japanese text, participants reported perceiving Japanese more than English (*only or mostly Japanese*: 31.7%; *only or mostly English*: 22.1%). When applied only to English text, the results were reversed (*only or mostly English*: 31.3%; *only or mostly Japanese*: 20.7%) (see Fig. 11). This suggests that it may be possible to influence which language text is perceived in a controlled BR experience. No correlation was found between the native language of participants, the majority of whom were native Japanese speakers, and the language which they perceived. This may indicate that low-level factors (such as movement) are more likely determinants of perceptual dominance than high-level factors (such as a more meaningful association with first language (L1) than second language (L2) information).

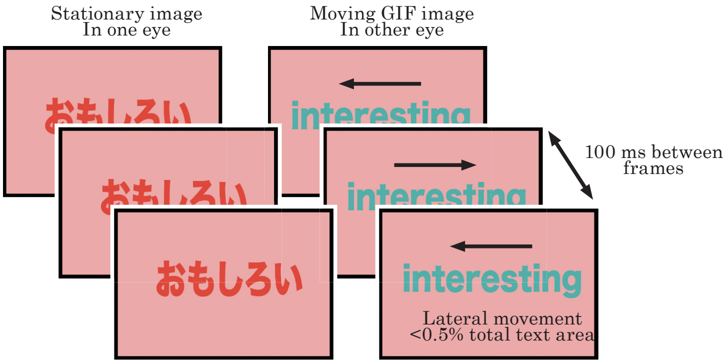

Figure 10: Moving GIF images were used to increase the stimulus strength of the target language via lateral movement.

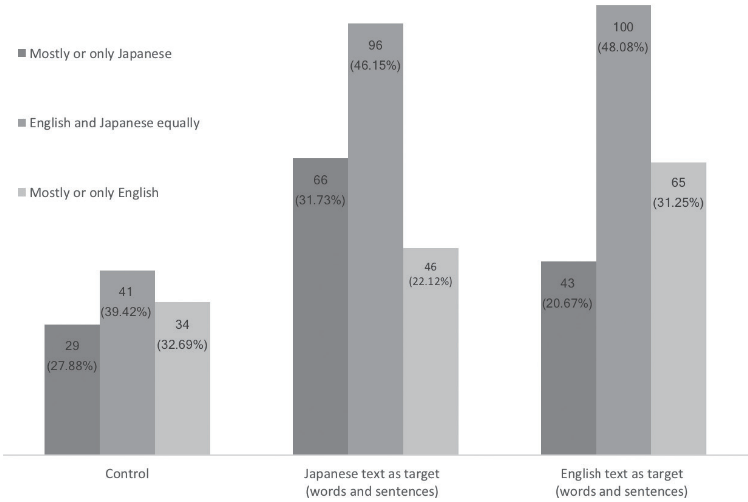

Figure 11: ***Reported dominant text language in control (no motion effect applied to either language text), Japanese text as target (motion effect only applied to Japanese language text), and English text as target (motion effect only applied to English language text) slides. By applying a motion effect to one of the two texts in a BR experience, the stimulus strength could be adjusted, resulting in the target language dominating conscious awareness.***

Discussion and possible applications

BR is a curious phenomenon and practical classroom applications may not be immediately obvious. However, it is worth remembering that binocular vision research is still in its infancy. As we learn more about how we process visual information, we may come to discover applications of the principles in seemingly unrelated areas, such as reading, viewing text, and language learning. As such, this paper can only act as a precursory investigation and offer suggestions.

3D glossing of bilingual texts is discussed in the preceding section of this paper as an experimental method of enriching the reading experience by information grouping at varying virtual depths. BR, too, offers possibilities in adding new dimensions to bilingual texts. By overlaying different language texts in the same location, and viewing them with anaglyph glasses, we are able to produce a BR experience in which our brain switches between the two. The duration each text is perceived depends on the relative stimulus strength of the text. By adjusting the strength of the texts, one can be made to dominate conscious awareness, while the other is suppressed. Digital reading applications for smartphones and tablets could allow the stimulus strengths to be precisely adjusted depending on the reader, allowing them to dynamically reveal or conceal the language they choose.

Vocabulary memorisation tools such as digital flip cards are increasingly popular amongst language learners (Khamees, 2016). The state-of-the art digital display of the ubiquitous smartphone has untapped potential as a means of creating highly engaging and interactive visual content. A BR experience which simultaneously displays text in the L1 and L2 of the user may provide, currently unknown, benefits to the language learner. Again, we should be careful not to over speculate and should consider the limitations of BR as an educational tool as discussed later in the paper.

Modularization of data

3D texts offer many more options than conventional texts for grouping text elements on a page: whereas a 2D text can only exploit proximity of text elements in the vertical and horizontal for grouping – the creation of relationships –, a 3D text offers depth as the third dimension for creating additional relationships.

It has been shown that teaching various content in a two-dimensional format (with tables, so-called “graphic organizers” or “matrix graphic organizers”) is superior to a pure linear representation of the same content, both for learning and retention of hierarchical and coordinate relations (Robinson & Kiewra, 1995). A 3D format has the potential for showing and teaching higher-dimensionally interconnected content.

Moreover, since our stationary environment is 3D, and our visual system creates mental 3D representations of this environment, we do have powerful cognitive resources for structuring, modularizing and selecting subsets of huge amounts of 3D – visual or other – data, which lay idle when we see, structure and decipher (i.e. read) conventional 2D texts. Thus, we hypothesize: the processing of the immense flood of data from the visual sense is only possible through systematic modularization and structuring. The same is true for reading when processing large amounts of characters. The development of reading comprehension and other language skills could be accelerated through the use of 3D vision and BR. Why? Binocular vision is specialized in constructing meaning(s) associated with perceived natural 3D scenes. The presentation of texts in 3D would support the modularization of textual elements, the conscious selection or filtering of elements on different depth levels, with different 3D orientations, etc. These distinct visual text elements would mean a well-structured visual modularization of information, which could correspond to corresponding mental modular structures. Focusing on one visual element –

e.g. one visual depth level – would then correspond to focusing attention on one mental module.

One possible example from language teaching are grammatical categories in the German language. Limitation to 2D make it difficult to show certain relationships, e.g. the identity of inflectional endings, which change according to several grammatical categories, but change identically for different words (see Fig. 12 & Fig. 13).

der	die	das	jeder	jede	jedes
des	der	des	jedes	jeder	jedes
dem	der	dem	jedem	jeder	jedem
den	die	das	jeden	jede	jedes

Figure 12: ***German is highly inflectional depending on various grammatical categories. Here: definite article "der ... das = the" (left side) and pronoun "jeder ... jedes = each" (right side), inflected identically according to gender (masculine, feminine, neuter) along the horizontal, and according to case (1st, 2nd, 3rd, 4th) along the vertical. Because the 2D tables are seperated, the corresponding identical endings cannot be grouped according to the proximity principle.***

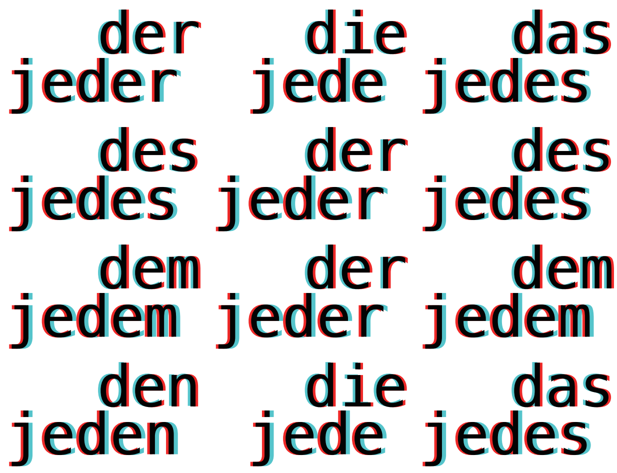

Figure 13: ***This anaglyph image shows the data from Fig. 12 in 3D; now the corresponding article-pronoun pairs can be placed closer to each other – along the 3rd dimension – and the identity of the endings can be seen at a glance. (Even without anaglyph glasses, the identity of the endings can easily be seen, since it is just the projection of the 3D table, however, the table looks somewhat crowded since all data is on the same plane.)***

Actual classroom application I: a grammar table in pseudo-3D

Anaglyph images have not been used in class. However, there are other cues for depth besides stereopsis, such as perspective (distant objects appear smaller), aerial perspective (distant colours appear fainter) and so forth.

The following pseudo-3D table has been used (Fig. 14) in class. It is actually a 2D table, but through using different font sizes and grey levels, different depth layers are mimicked; the table looks less crowded compared to Fig. 13 (without anaglyph glasses). In other words: the grouping principle of similarity (here: same font size, same grey level, similar characters) helps to see three 2D tables in one.

	Singular			Plural	
	m	f	n	pl	
1	de**R** gute Mann ein gute**R** Mann gute**R** Mann	di**E** gute Frau ein**E** gute Frau gut**E** Frau	da**S** gute Kind ein gute**S** Kind gute**S** Kind	di**E** guten Leute (kein**E** guten Leute)* gut**E** Leute	def. art. indef. art. no article.
4	de**N** guten Mann eine**N** guten Mann gute**N** Mann	di**E** gute Frau ein**E** gute Frau gut**E** Frau	da**S** gute Kind ein gute**S** Kind gute**S** Kind	di**E** guten Leute (kein**E** guten Leute)* gut**E** Leute	def. art. indef. art. no article.
3	de**M** guten Mann eine**M** guten Mann gute**M** Mann	de**R** guten Frau eine**R** guten Frau gute**R** Frau	de**M** guten Kind eine**M** guten Kind gute**M** Kind	de**N** guten Leuten (keine**N** guten Leuten)* guten Leute**N**	def. art. indef. art. no article.
2	de**S** guten Mannes eine**S** guten Mannes guten Manne**S**	de**R** guten Frau eine**R** guten Frau gute**R** Frau	de**S** guten Kindes eine**S** guten Kindes guten Kinde**S**	de**R** guten Leute (keine**R** guten Leute)* gute**R** Leute	def. art. indef. art. no article.

*There is no indefininte article in plural, but "kein_" (no), the negation of "ein" (no), is formed as indicated (like "mein_", ...)

Figure 14: ***Grammar table in pseudo-3D used in a handout: distance from the reader (depth) is mimicked by font size + grey level so that the three lines in each cell might be seen at three different depths. This gives a slight 3D-impression and makes the table appear less crowded. As in Fig. 13, this is an inflection table, here for noun phrases and in particular for the adjectives ("gut_" = "good"), here gender/number (m-f-n-pl) along the horizontal, case (1-4-3-2) along the vertical, and (definite-indefinite-no) article along the depth dimension.***

Actual classroom application II: text with reduced information (in normal view)

The following example with reduced information, a gap-filling puzzle, has been used in class. It involves a text with reduced and – hidden for the normal user – increased reduncancy. The goal was given as follows:

PUZZLE (optional): The missing letters give a sentence: First solve the left side (about half the letters), then solve the right side (the rest). While solving one side, try not look on the other side! 😀
The solution: D a _ ____ "__ ___ __ ___ ____"___ ____ ____ ___

The actual text was supplied in a two-column format online (Fig. 15).

Klaus: Hallo Thomas. | Thomas:
Hallo Klaus. | Was machst du? |
Nichts. Ich habe Hunger. Können wir
etwa_ essen? | Ja, gerne! Ich habe
auch Hunger. Gehen wir in die Mensa?
| Nein. Ich mag die Mensa nicht so
gern. | Was möchtest du essen? | Ich
möchte etwas Scharfes essen. |
So_len w_r ungarisch _ssen? | Was
sagst du? Ungarisch? _as ist s_hr
scharf. Kann_t du ungar_sch es_en? |
Ja, ich kann heu_e mal etwas _nderes
tu_. | OK. _ann gehen wir! | Mit d_m
Fahr_ad? | Ja, fahren wir mit dem
Rad! | Alles klar! Auf _um Essen!

Klaus: Hallo Thomas . | Thomas:
Hallo Klaus . | Was machst du? |
Nichts . Ich habe Hung_r . Können w_r
e_was essen ? | Ja , gerne ! _ch habe
auch Hunger. Gehen wir in die Men_a ?
| Nein . Ich mag die Mensa nich_ so
gern. | W_s möchtest d_ essen? | Ich
mö_hte etwas Sc_arfes e_sen. |
Soll_n wir ungarisc_ essen? | Was
sagst du? Unga_isch ? Das ist sehr
scharf. Kannst du un_arisch essen? |
Ja , ich kann he_te mal etwas anderes
_un. | OK . Dann gehen wir! | Mit dem
Fahrrad ? | Ja , fahren wir mit dem
Rad ! | Alles klar ! Auf zum Essen !

Figure 15: ***Text with reduced redundancy (puzzle) in normal view, which can also be seen as a free-viewing stereogram (left & right column almost identical), making the missing letters (solution) visible: the gaps & the missing letters from the opposite side create BR stimuli in which the missing letters (left or left) are dominant, hence the original text can be read in stereoscopic view. In this view, the image even turns into a text with enhanced information, since words with sentence stress appear elevated (e.g. "Thomas", "Klaus", "machst", "Nichts", "Hunger" ...). (3D features not visible with anaglyph glasses, since this is a free-viewing stereogram. Distance of half images should not exceed ca. 60 mm for parallel view.)***

There are two characteristics to the text and to the related activities when seen without 3D effect, yielding a text with reduced redundancy (gap-filling task).

Repetition task: simple gap-filling (one-letter gaps).

- The missing letters put together form a meaningful sentence: "Das Lied 'Es ist an der Zeit' ist auch sehr gut."="The song 'Es ist an der Zeit' is also very good. " → The repetition task is turned into a puzzle.
- The right and left columns are identical, except the gap positions: here, the missing letters are on the corresponding positions on the other side. → Solvable even by lower level students (through a quick saccade to the other side).

Two test readers perceived increased redundancy / enhanced information (in stereo view)

However, there is more to the puzzle above, due to the potential of the two texts to combine to a 3D experience:

1. Seen as a free-viewing stereogram (parallel view or cross-eyed view), the complete text can be read: missing letters appear as half images in the fused image, and each missing letter on one half image competes with an underscore character on the other half image (BR stimuli). In general, the missing letter will be dominant, since it supplies the more interesting content.
2. Seen as a stereogram (parallel view), stressed words/names appear elevated relative to the rest of the text (30 elevated words mark 30 keywords).

Two test readers, TR1 & TR2 (both of whom were language learners capable of perceiving stereograms in free-viewing mode, TR1 younger than 25, TR2 older than 50, participation by informed consent) read aloud the text in Fig. 15 repeatedly several times, cycling through four conditions: (A) monoscopic (=normal) view of the left text, (B) monoscopic (=normal) view of the right text, (C) stereoscopic view with normal intonation & (D) stereoscopic view but instructed to exaggerate the stresses of the elevated keywords. The last cycle was done six or seven days after the first cycle.

Results: both readers realized their personal maximum of correctly stressed keywords in condition (D): TR1 achieved 27, TR2 all 30 stresses (min./max. for A, B, C & D: TR1: 21/25, 20/26, 23/26 & 25/27. TR2: 23/25, 19/24, 22/25 & 26/30). Thus, recognizing and stressing keywords marked by depth in a 3D text is possible in a read-aloud task. The overall reading time gave no clear preference of conditions: the final times of the stereoscopic modes were the overall shortest times for TR2 (max/min of all reading times: 103 s/63 s), but not for TR1 (max/min of all reading times: 70 s/57 s): TR1 might finally have been able to fill the gaps from memory without looking at the other side, nullifying the advantage of stereo view for gap-filling activities. The difference of the max. reading times between TR1 (70 s) and TR2 (103 s) might stem from a difference in text familiarity: TR1 knew the text from class (though he had not read it aloud nor stereoscopically before), whereas TR2, a false beginner, had no knowledge of the text prior to the experiment.

Conclusion: reading 3D texts while extracting and interpreting 3D features from words is principally possible, but further experiments are needed to explore the influence on reading speed.

Limitations of binocular vision for language learning

Stereoblindness is a condition where stereoscopic depth cannot be perceived using stereopsis. It can result from both physiological and neurological factors and is estimated to affect around 7% of people (Chopin et al., 2019). Stereoblindness exists at various degrees, thus it is difficult to estimate the extent to which the condition will impact the binocular experience of 3D. However, we may assume that many stereoblind language learners would not benefit from simulated 3D. This problem may be overcome to some degree with the use of pseudo-3D images (see Fig. 14), which also employ the principles of visual grouping.

Ocular dominance is the tendency to rely on one eye more than the other (Khan & Crawford, 2001). Most of us have a preferred eye to some degree. Usually this does not cause any problems in everyday life, but highly profound cases can create challenges. In terms of the BR experience, it is unsurprising to learn that the information in the dominant eye tends to be that which we perceive. Those with a strong ocular dominance are more likely to have an unbalanced experience, with one image or text dominating for longer than the other. While this may, to a certain degree, be rectified by calibrating the stimuli to match the vision of the individual, it should be acknowledged that it does present added complications.

Overreliance on unnecessary technology is understandably a cause for concern in education today. While the learning benefits offered by laptops, smartphones, and tablets are clear, they also present educators with several new, and not yet fully understood, challenges in regard to learners' deteriorating linguistic and social competencies (Alhumaid, 2019). Where developments in educational tech may lead is impossible to know, but it is certain that new solutions will be found to current problems. As knowledge about our perceptual systems grows in tandem with technological advancements, elegant assistive technology which exploits our innate biology may soon be realized. We can only hope that these solutions will add something to the learning experience without distracting students from it.

Conclusion

The way by which we derive meaning from visual information is a complex process of which our understanding is constantly evolving. In this paper, we suggested the use of binocular vision, in particular 3D vision and BR for foreign language education. Possible applications include glossing of foreign language texts, vocabulary memorization and 3D grammar tables. The paper also suggests technology-free applications (such as pseudo-3D tables) for the inclusive classroom and to account for students with varying levels of ocular dominance, stereoblindness, etc. We believe that a balanced usage of both technological innovation and our innate physiology yields the optimum potential for language education. Often overwhelmed by technological advances, we forget too easily, that our cognitive system also evolves in response to said technology and the way in which we use it. Our forthcoming research will strive to elaborate on solutions which exploit the intricate and everchanging relationship between binocular vision, specifically 3D and BR, and learning.

Note on Figures and citation of student comment:

All images were made by the authors. Except:

Figure 1: [PSM V21 D049 Wheatstone stereoscope 1.jpg] (2010). Retrieved from https://commons.wikimedia.org/wiki/File:PSM_V21_D049_Wheatstone_stereoscope_1.jpg#filelinks

Figure 17: This figure was created by a student and is used with informed consent.

Appendix: The student comment on the impact of PW is cited with consent.

References

Alais, D., & Blake, R. (2005). *Binocular rivalry.* Cambridge, MA: MIT Press.

Alhumaid, K. (2019). Four Ways Technology Has Negatively Changed Education. *Journal of Educational and Social Research, 9*(4). doi:10.36941/jesr-2019-0002

Bessemans, A., Renckens, M., Bormans, K., Nuyts, E., & Larson, K. (2019). Visual prosody supports reading aloud expressively. *Visible Language*, *53*(3), 28-49.

Carroll, L. (2007). *Alice's Abenteuer im Wunderland [Alice's adventures in wonderland*]. Urbana, Illinois: Project Gutenberg. Retrieved from http://www.gutenberg.org/cache/epub/19778/pg19778.txt

Chopin, A., Bavelier, D., Levi, D. (2019). The prevalence and diagnosis of "stereoblindness": A best evidence synthesis. *Journal of Vision,19*(10): 262b. doi:https://doi.org/10.1167/19.10.262b.

Fox, R. (1978). Binocularity and Stereopsis in the Evolution of Vertebrate

Vision. *Frontiers in Visual Science Springer Series in Optical Sciences,* 316-327. doi:10.1007/978-3-540-35397-3_32

Hall, W. (2020). How to read Two Books at Once: Binocular Rivalry and Foreign Language Learning. *Mindbrained Journal*, 2.

Hiragana Times. (2020). *Hiragana Times: Japanese and English Bilingual Magazine.* December 2020 No. 410.

Hoffman, D. (2000). *Visual Intelligence.* New York: Norton.

Howard. I., Rogers B. (1995). *Binocular vision and stereopsis.* New York: Oxford University Press.

Khamees, Khalid. (2016). An Evaluative Study of Memorization as a Strategy for Learning English. *International Journal of English Linguistics.* 6. 248. 10.5539/ijel.v6n4p248.

Khan, A., & Crawford, J. (2001). Ocular dominance reverses as a function of horizontal gaze angle. *Vision Research, 41*, 1743-1748.

Levelt, W. J. (1968). *On binocular rivalry.* Paris: Mouton.

Nation I.S.P. (2008). *Teaching Vocabulary. Strategies and Techniques.* Boston: Heinle.

Nation I.S.P. (1990). *Teaching and Learning Vocabulary.* Boston: Heinle & Heinle.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interest, 2*(2), 31-74. doi:10.1111/1529-1006.00004

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of educational psychology*, *87*(3), 455.

Rude, M. (2016). Prosodic Writing shows L2 learners intonation by 3D letter shapes: state, results, and attempts to increase 3D perception. *Nagoya University, Studies in Language and Culture*, 103-120.

Salem, E. B. (2006). The influence of electronic glosses on word retention and reading comprehension with Spanish language learners. *Unpublished dissertation submitted to the Faculty of the Graduate School of Education of the University of Kansas.*

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* University of Illinois Press.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological bulletin*, *138*(6), 1172.

Wilson, H. R., & Julesz, B. (1971). Foundations of Cyclopean Perception. *Chicago Review, 24*(3), 146. doi:10.2307/25294744

Wolfe, J. M., Kluender, K. R., & Levi, D. M. (2020). *Sensation & perception.* New York: Sinauer Associates: Oxford University Press.

Zeki, S. (2003). *Inner vision: An exploration of art and the brain.* Oxford University Press.

Appendix: Texts with enhanced phonetic information

Stresses belong to phonetics, a field rich with potential applications for binocular vision. The two images here – a pseudo-3D image and a free-viewing stereogram – show texts with additional phonetic information. Prosodic Writing (PW) is a visualization of spoken language in three dimensions (Rude 2016): the *vertical* displacement expresses *pitch*, the *horizontal* dimension *time* and the *depth* dimension *loudness* (expressed through size/perspective). Pitch, time and loudness, however, mainly constitute intonation, rhythm and stress, i.e. prosody. Thus, PW visualizes prosody and can help language learners to grasp intonation, rhythm and stress. Fig. 16, a pseudo-3D image, depicts a handwritten example of PW.

PW is rather intuitive and can even evoke bodily reactions: "I felt bit ashamed that I can't help moving my head up and down when reading PW sentences." (Reported by one out of ca. 50 students, who seemed to be surprised about the own reaction but who also confirmed later that this reaction was non-distracting.)

Figure 16: ***Shown are two questions from a handout for a test preparation ("How are you?" & "What music do you like?"), a text with enhanced information: prosodic cues (stress and high pitch accents on "geht" (=goes) and "Musik" (=music), secondary stress on "welche" (=which/what)"). The students received 40 such questions accompanied by an MP3 file.***

For a related visualization (height of characters above baseline express pitch, font sizes express loudness, etc.) it was shown that trained children could significantly improve their expressive reading skill when working with it (Bessemans et al., 2019).

Thus, given the biological relevance of depth perception (distance to predator or prey), 3D visualizations might further increase the intuitiveness of PW. As a first feedback from the uses in class (e.g. pseudo-3D Prosodic Writing, puzzle stereogram, etc.), one student came up and supplied a version of stereoscopic Prosodic Writing in his learner diary, made by the free and open source 3D creation software package "Blender", and shown in Fig. 17.

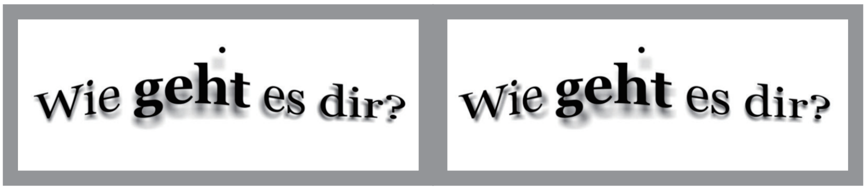

Figure 17: Stereogram showing the elevated word "geht" on grounded shadows (base plane). Already in normal view, the upcurving letter string on "geht" suggests a high pitch accent, typical for sentence stress in German language (additional prominence by boldness, suggesting maximum loudness that often accompanies sentence stress). 3D elevation, visible in stereoscopic parallel view and then maximal on "geht", could further raise the intuitiveness of PW (stereogram by Ryosuke Takahashi, computer science student at Tsukuba University; frame added by authors to simplify the fusion of both half images. Distance of half images should not exceed ca. 60 mm for parallel view).