

# Advanced Colour Terms

## - Beyond Human Vocabularies: Artificial Language and the Perception of Colour

Hanna Hofmann<sup>1</sup>

<sup>1</sup> University of Applied Arts Vienna

### Abstract

*Advanced Colour Terms* is an artistic research installation that explores the relationship between linguistic structure, visual perception, and machine learning through the artistic generation of new colour terms. Based on the theory of *Basic Colour Terms* (Berlin & Kay, 1969), the project examines whether artificial systems can imitate or rethink the linguistic encoding of colour as shaped by human perceptual processes. A recurrent neural network was trained using over 300 phonetic transcriptions of colour terms from around the world, enabling the creation of new colour descriptors based on semantic categories. These terms were vocalised using a modified multilingual text-to-speech system (IMS Toucan), which neutralised linguistic bias by averaging language embeddings and blending synthetic voices. The resulting speech was synchronised with the corresponding light output in the final installation to create a continuous loop between the generated language and the perceived colour—thus asking whether artificial language, born from human data but independent of a biological systems, can meaningfully replicate or transform our experience of colour.

### 1. Introduction:

Advanced Colour Terms is an artistic installation that investigates the linguistic and perceptual boundaries of colour naming. Two industrial pipes emerge from the floor, establishing a sensory dialogue between light and sound. One pipe emits shifting fields of coloured light, while the other produces unsettling, machine-generated speech.

At the centre of this interaction lies a machine learning model trained on over 300 colour terms collected from languages across the world. Drawing on this linguistic archive, the model generates entirely new colour descriptors, speculating on how a non-human intelligence might expand, disrupt, or reinvent the ways we categorise and name colours. In doing so, it also seeks to identify underlying patterns in how humans have historically encoded colour perception into language—patterns famously described by Berlin and Kay's theory of Basic Colour Terms [1].

By combining methods from computational linguistics, perceptual psychology, and machine learning, the work explores how language not only reflects but actively shapes sensory experience. By synthesising speculative colour names and sonifying them using a multilingual trained synthetic voice system, the installation poses questions about linguistic loss, possible rebirth through machine creativity, and the inherently unstable nature of perception itself.

### 2. Theoretical Background

#### Basic Colour Terms and Universality

Research on colour naming has historically been shaped by the work of Berlin and Kay (1969) [2], who proposed that all languages draw from a universal set of eleven basic colour categories: black, white, red, green, yellow, blue, brown, pink, orange, purple, and grey. They argued that if a language has fewer terms, these are acquired in a predictable evolutionary sequence—from dark/light contrasts to more specific hues.

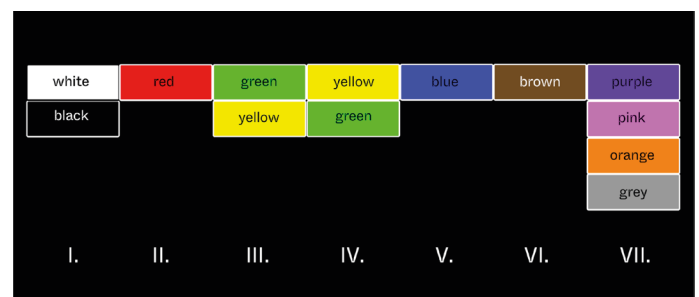


Fig.1: Basic Colour Terms

Kay and McDaniel later expanded this view, arguing that these universals arise not from linguistic arbitrariness (as the Sapir-Whorf hypothesis would suggest [3]), but from the structure of human perception itself.

According to their research, the way humans divide the continuous colour spectrum into distinct linguistic categories is reflective of the innate properties of the human visual system. These

properties include the sensitivity of the retina's cone cells to certain wavelengths of light, and the organisation of colour processing in the brain. From this standpoint, language does not impose arbitrary boundaries on a neutral perceptual field, but rather emerges in response to perceptual salience. Certain regions of the colour spectrum are more distinct to humans, and these regions tend to form the basis of universal colour categories across languages.

More recent computational models, such as the Information Bottleneck Principle, frame colour naming as a problem of efficient communication [4]. They propose that languages evolve to balance precision—the ability to distinguish many colours—and efficiency—using as few terms as possible. This leads to the compression of perceptual continua, like colour, into optimally spaced categories. Such models suggest that colour lexicons are shaped not by chance, but by universal pressures of perception, memory, and communicative need, positioning language as a dynamic system co-evolving with human sensory experience.

### 3. Methodology

Starting from Berlin and Kay's categories, I created a dataset by collecting over 300 colour terms in multiple languages from Omniglot's multilingual colour resource. Each term was manually transliterated into the International Phonetic Alphabet (IPA) to standardize pronunciation across linguistic diversity. This step was essential to prepare the data for neural language modeling.

Machine Learning Model: Generating New Colour Terms

#### Machine Learning Model: Generating New Colour Terms

To expand on the known inventory of colour terms, I developed a machine learning model that generates entirely new, speculative colour names. This model is based on a type of neural network called a Recurrent Neural Network (RNN), which is designed to handle sequences, such as strings of letters or sounds.

The model takes two kinds of information as input:

- A colour category (for example, blue, red, or yellow), which provides semantic guidance.
- The spelling of existing colour terms, written in the International Phonetic Alphabet (IPA), which provides examples of how colour words sound across different languages.

By learning from both the meaning (the colour category) and the form (the phonetic spelling), the model can generate new terms that sound like they could belong to real languages, but are actually entirely invented and merged forms of

```
color_input = Input(shape=(1,),
name='color_input')
char_input = Input(shape=(max_len-
1,), name='char_input')
color_embed = Flatten()(Embed-
ding(input_dim=num_colors, output_
dim=16)(color_input))
color_embed = RepeatVector(max_
len-1)(color_embed)
char_embed = Embedding(input_
dim=vocab_size, output_dim=64)
(char_input)
combined = Concatenate()([char_em-
bed, color_embed])
lstm_out = LSTM(128, return_se-
quences = True, dropout=0.2)(com-
bined)
output = Dense(vocab_size, activa-
tion='softmax')(lstm_out)
```

Fig.2: Example Model Snippet

already existing ones.

To maximize the creative variability of the output, I employed temperature sampling and a diversity scoring function that prioritizes uniqueness in the generated terms.

#### Sonic Rendering with IMS Toucan

After generating new colour terms, I used *IMS Toucan*, a state-of-the-art multilingual text-to-speech (TTS) system capable of synthesising speech in over 7,000 languages, to vocalise these speculative linguistic forms [5].

In the conceptual development of Advanced Colour Terms, I sought to create a synthetic voice that would resist the linguistic and cultural biases embedded in human speech. My strategy was to vocalise the machine-generated colour terms in a way that felt artificially neutral—a voice that could not be easily located within any specific language, accent, or speaker identity. Inspired by this, I initially experimented with layering multiple synthetic voices in audio software, attempting to blur individual speech characteristics through manual mixing. However, this process was labour-intensive and technically limited.

In conversation with Florian Lux, one of the lead developers of IMS Toucan, I shared this conceptual strategy. Motivated by the artistic challenge,

Florian extended the Toucan codebase to support multi-layered, accent-agnostic voice synthesis directly within the TTS system itself. This collaboration resulted in two key technical modifications:

### **1. Neutralised Language Embedding**

Florian Lux modified the system to average multiple language representations, effectively removing the linguistic characteristics that typically define the pronunciation in a single language. The selected languages for averaging were chosen based on their wide geographic and phonetic diversity, ensuring the resulting voice did not retain any recognisable accent or language-specific bias. This produced a vocalisation that sits outside of human linguistic norms—an abstract, machine-like voice that reflects the speculative, non-human origin of the generated terms.

### **2. Artificial Multi-Voice Layering**

In addition to neutralising the linguistic layer, the system generates multiple artificial speaker identities using fixed latent seeds (3 to 7). These seeds correspond to pre-computed, non-human speaker profiles designed to sound distinct. For each generated colour term, the system produces multiple vocalisations using these seeds and blends them into a single, layered voice. Importantly, the pitch, energy, and timing are synchronised across all voices — predicted once for the first voice and then reused — creating the effect of a polyphonic, perfectly aligned chorus. I have experimented with this technique in Ableton for the original concept of layering voices by hand, but this technique even more so enriches the acoustic texture while also reinforcing the artificial nature of the speech, distancing it from any recognisable human speaker.

#### **Personal Embodiment Layer**

Finally, layered on top of these synthetic voices is a cloned version of my own voice. This personal layer introduces a subtle human presence into the otherwise artificial soundscape, situating my authorship and body within the work. It creates a tension between machine-generated abstraction and personal embodiment, reflecting the collaborative nature of this human-machine system.

Together, these techniques result in a complex vocal texture that merges linguistic diversity, machine learning abstraction, and human interpretation. The result is a voice that is at once recognisably artificial and eerily cohesive—situating the audience within a sensory space where language, technology, and perception co-produce

new forms of meaning.

## **4. Installation Concept**

Two industrial pipes create a sensory loop:

One emits shifting coloured light, corresponding to a generated term.

The other releases the speculative colour term as layered synthetic speech.

This real-time interplay between light and sound transforms linguistic data into a multi-sensory feedback system, inviting the audience to consider the artificial and natural limits of perception, language loss, and technological re-creation.

The installation is designed as a sensory feedback system consisting of two industrial pipes emerging from the floor to create an ongoing exchange of light and sound. One pipe emits slowly shifting fields of coloured light, each of which corresponds to a newly generated speculative colour term. The same term is voiced aloud from the other pipe through a layered artificial speech synthesis system that blends multiple machine-generated voices into a unified sound.

This invites the audience to reflect on the entangled relationship between perception and language—how our visual system has historically shaped the way we name colours, and how those names, in turn, reinforce the way we see. As light and sound play out in parallel, the installation creates a loop in which artificial language and visual perception shape each other continuously.

## **5. Discussion**

Advanced Colour Terms builds on the foundational theory of Basic Colour Terms, which proposes that all languages develop colour vocabularies following a universal sequence shaped by human visual perception. The work uses this principle as both method and challenge— a machine learning model is trained using over 300 IPA-transcribed colour terms from languages spoken around the world. The model then generates new descriptors to explore the extent to which a computational system can mimic or reimagine the linguistic structures tied to human sensory experience. The installation reverses the perceptual-linguistic relationship proposed by Berlin and Kay by feeding colour categories and phonetic patterns into the model. This approach starts not from vision, but from language, exploring how a machine without sensory experience might invent its own colour vocabulary.

In doing so, the work raises broader questions: Can perception be encoded purely through language? Can a machine, trained on the outcomes of perceptual experience, reverse-engineer this experience itself? And if the link between colour and language is rooted in biology, what does it mean when a non-biological system begins to generate its own colour vocabulary?

*Advanced Colour Terms* explores the extent to which machines can engage with and reconfigure sensory categories, and the implications of language emerging independently of the biological body. It questions whether these invented terms, detached from human embodiment yet shaped by human data, can still elicit a sense of colour—whether meaning and sensation can survive, or even transform, through artificial language:

What is lost, reshaped, or newly possible when the act of naming shifts beyond our human experience?

---

## 6. References

- [1] Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- [2] Kay, P., & McDaniel, C. K. (1978). *The linguistic significance of the meanings of basic color terms*. *Language*, 54(3), 610–646.
- [3] Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.
- [4] Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). *Efficient compression in color naming and its evolution*. *Proceedings of the National Academy of Sciences*, 115(31), 7937–7942. <https://doi.org/10.1073/pnas.1722022115>
- [5] Lux, F., Deng, Y., & Tits, N. (2024). *Meta learning text-to-speech synthesis in over 7000 languages*. arXiv preprint arXiv:2406.06403. <https://doi.org/10.48550/arXiv.2406.06403>