



Factors affecting the percept of *Yanny* v. *Laurel* (or *mixed*): Insights from a large-scale study on Swiss German listeners

Adrian Leemann¹, Péter Jeszenszky¹, Carina Steiner¹, Corinne Lanthemann¹

¹Center for the Study of Language and Society, University of Bern
adrian.leemann@unibe.ch, peter.jeszenszky@unibe.ch
carina.steiner@unibe.ch, corinne.lanthemann@outlook.com

Abstract

In May 2018, *Yanny* v. *Laurel* went viral: when listening to the same audio clip, some people claimed to hear only *Yanny*, others insisted it must be *Laurel*, and some had a mixed percept. Phoneticians have identified the acoustic features which caused this perceptual ambiguity, but we still know little about the factors affecting individuals' perception of the illusion. We conducted a controlled study with 974 Swiss German listeners, balanced for age, gender, and regional origin. Overall, nearly two thirds heard *Yanny*, one quarter *Laurel*, and about 12% had a mixed percept. We found age, gender, and electronic device to play a significant role: younger, female, and laptop-using participants demonstrated higher proportions of *Yanny* responses. These findings contribute to the growing body of research on polyperceivable words.

Index Terms: speech perception, auditory illusion, Swiss German

1. Introduction

In May of 2018, a recording of a single word caused widespread debate: we are, of course, talking about the ominous-sounding file with a percept of either *Yanny* or *Laurel* (or something else altogether). The origin of the clip makes for a good story: a high school student –Katie Hetzel – was looking for the pronunciation of the name *Laurel* and checked the site www.vocabulary.com. The student saved the recording and, upon replaying it, noticed that she no longer heard *Laurel*, but *Yanny*. She mentioned this to a friend who uploaded it on social media and it went viral from there (the original recording can be found here: <https://twitter.com/CloeCouture/status/996218489831473152>). Listeners were bewildered and astonished by the illusion. Interestingly, many listeners claimed they could only access a single interpretation of the stimulus, suggesting that the percept is very stable – and probably contributing to the widespread debate on the sound file. In essence, the world had discovered an acoustic equivalent to #TheDress [1].

The reason for the robust perceptual divide becomes clear when we look at the acoustic signal in a bit more detail. Figure 1 shows the spectrogram of the original signal (center) and articulations of *Laurel* (left) and *Yanny* (right). The latter two were created through the splicing of phonemes taken from other words articulated by the male who recorded 'Laurel' for vocabulary.com (cf. Figure 1, adopted from [2]).

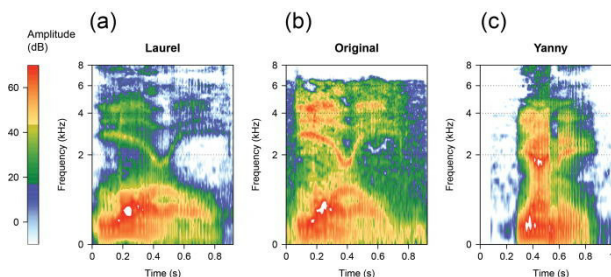


Figure 1: Comparison of the original (center) with spliced articulations of *Laurel* (left) and *Yanny* (right) (cf. [2]).

The spectrographic representations in Figure 1 show that the two spliced recordings of *Laurel* (left) and *Yanny* (right) are particularly different in one respect: *Laurel* shows high acoustic energy (red) in the lower frequency domain, while *Yanny* shows high acoustic energy in both the lower and upper (~2K Hz) frequency domains. The ambiguous, original *Yanny* v. *Laurel* recording (center) displays a bit of both: here, higher frequencies (> 2K Hz) are particularly strong, especially in the first 400ms of the signal (shown as the downwards trajectory at around 3K Hz from 0 to c. 0.4s). It is unclear why that is, perhaps because of the (poor quality) device the student had used to record the alleged articulation of *Laurel*.

It has been argued that this pronounced energy is what triggers the ambiguity: some listeners interpret it as the F3 of /ɔ/ and hear *Laurel*, while others perceive it as the F2 of /æ/ and hear *Yanny* [2]. When the file is perceived as *Yanny*, the /æ/ sounds raised and fronted, typical of American English prenasal TRAP tensing [3]. Because of the acoustic ambiguity in the signal, our minds have several degrees of freedom in how the stimulus can be interpreted, which then manifests in different proportions of percepts: for example, using a crowdsourcing approach to survey 532 listeners, [2] reports 60% *Laurel*, 30% *Yanny*, and 10% mixed percepts.

Several studies have conducted signal manipulations to explore whether listeners can be induced to hear one percept or the other: [4] for instance, exposed participants to either low-pass or high-pass filtered versions of the original ambiguous stimulus. The low-pass filtered version triggered more *Laurel* percepts and the high-pass filtered version more *Yanny* percepts – which was expected, as *Laurel* has greater energy in the lower frequency regions and *Yanny* contains a high amount of energy in both lower and higher regions. [2] conducted a more nuanced attempt to induce one or the other percept by modulating lower and higher frequencies in a 7-step continuum. Figure 2 illustrates this continuum, which gradually triggered more

Yanny responses the more high-pass filtering that was applied – shown from left to right in Figure 2.

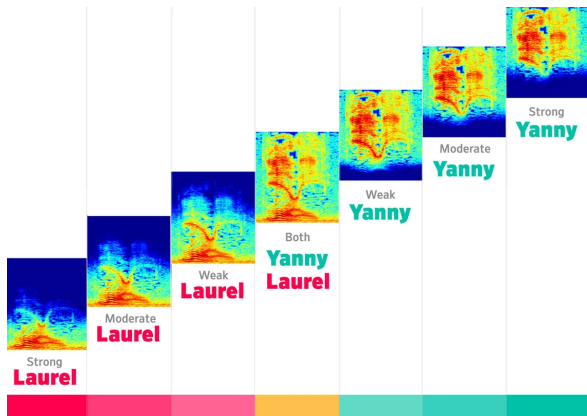


Figure 2: 7-step phonetic continuum triggering a percept from *Laurel* to *Yanny* (adopted from [5]).

Just as a mobile phone screen’s level of brightness played a role in the perception of #TheDress [1], several social, physiological, and hardware-related factors have been shown to affect the perception of *Yanny* v. *Laurel*. Listening equipment, experience, listeners’ first language, context, and gender all seem to play a role. [4] showed that (self-reported) listening equipment affects listener responses: over-ear and in-ear headphones triggered a higher proportion of *Laurel* percepts than laptop speakers did. Secondly, experience – operationalized via ‘musicality’ – appears to be crucial: musically-inclined listeners heard a higher proportion of *Laurel* [4]. In terms of listener language, [4] found that L1 Anglophone listeners were more likely to perceive a higher proportion of *Laurel* compared to L1 Francophones. Finally, recent evidence suggests that context plays a role in priming one percept or the other: [2] played audio precursors before the *Yanny* v. *Laurel* stimulus; these precursors were a sequence of numbers that were either low-pass or high-pass filtered. If the precursor was low-pass filtered, the subsequent percept tended toward *Yanny* because listeners paid more attention to higher frequencies in the target stimulus (and *Yanny* has more acoustic energy in the higher frequencies than *Laurel*); vice-versa for the high-pass filtered precursors. Finally, existing research also recognizes the role of listener gender in the percept of *Yanny* or *Laurel*: [4] reported men to perceive a higher proportion of *Laurel*.

The generalizability of this previous research is not entirely unproblematic, however. All of these studies have relied on crowd-sourced, self-reported data, which is notoriously noisy. [2], for example, did not collect data on the 532 participants’ genders, ages, first languages, or listening equipment, even though these factors have been reported to affect perception. The [4] study is somewhat unsatisfactory because they appear to have used a predominantly binary forced choice paradigm on their 289 participants. While they did consider what listening equipment was used by the participants, they did not collect information on the electronic devices used, which could possibly play a role since devices have different frequency responses [6].

The purpose of the present study, then, is to explore the percept of the *Yanny* v. *Laurel* stimulus in a more controlled fashion and with a balanced listener sample. Given the previous literature, going into the study we expected to find an effect of

gender, electronic device, and musicality. We further predicted a potential effect of region, given substantial variation in vowel quality, quantity, and phonotactics across Swiss German dialects (cf. [7]) – the assumption being that listeners have a perceptual preference for phoneme sequences that are more frequent in their dialect [8]. We further expected competence in English to play a role, given that competence in an L2 can affect discrimination ability in that foreign language [9]. Finally, we suspected age to play a role: as mentioned, the words *Yanny* and *Laurel* differ particularly in the higher frequency regions; these regions are likely to be attenuated in the hearing of elderly speakers, thus triggering higher proportions of *Laurel* percepts in older listeners.

2. Methods

2.1 Material

The original *Yanny* v. *Laurel* stimulus was retrieved from here <https://twitter.com/CloeCouture/status/996218489831473152> and was normalized in amplitude to 69dB in Praat [10].

2.2 Listeners

One thousand listeners from 125 localities across German-speaking Switzerland participated in this study. They came from eight major regions, covering all major areas in German-speaking Switzerland (Cantons of Bern, Zurich, Grisons, Aargau, Northwestern Switzerland, Northeastern Switzerland, Fribourg, Valais & Ticino, as well as Central Switzerland – survey site selection is described in [11]). Figure 3 shows the survey sites.



Figure 3: Survey sites from SDATS [10].

The sample consisted of 500 males and 500 females who took part in the SDATS project [12]. Half of the speakers belonged to a younger cohort and half to an older cohort (younger cohort mean age: 26, SD=4.8; older cohort mean age 70, SD=6.6). 55% claimed to be musical (i.e., sang or played an instrument in their free time or professionally), 36% had a tertiary education degree, and 37% used English on a regular basis and were relatively proficient (mean 3.3 on a 7-point rating scale). 23% of the participants had impaired hearing. Questions on hearing impairment and musicality were not mandatory, however, and 26 participants did not provide this information. For this reason, they were excluded from the sample, which meant the final dataset consisted of N=974 participants.

2.3 Procedures

All listeners took part in the SDATS dialect survey [12], which, amongst other tasks, asked about the perception of *Yanny* v. *Laurel*. Due to the ongoing COVID-19 pandemic, data collection occurred remotely for most listeners (76.2% of the full dataset). Remote participants were generally located at their homes and linked with us on Zoom as we guided them through the linguistic tasks; they recorded their responses remotely on their smartphones using a bespoke app (cf. [13] for a description of the recording set-up). Figure 4 demonstrates the prompt enabling the collection of responses to the *Yanny* v. *Laurel* stimulus.



Figure 4: Screenshot of the instructions (left) and *Yanny* v. *Laurel* audio prompt (right) shown to the participants.

Listeners were instructed to carefully listen to the supervisor’s explanations (cf. Figure 4, left panel) before clicking ‘play’ on the next slide (cf. Figure 4, right panel). Output volume was set at a comfortable level from a previous task. Listeners were instructed to play the file only once and then record their answer. 54% of the listeners heard the stimulus on a device running iOS, 38% Android, and 8% (N=79) on laptops (all Macs with varying specifications), using *SpeechRecorder* [14]. The 79 laptop-based participants completed the survey in a face-to-face setting. Of the Android devices, 23% were Samsung, 7% Huawei, and 70% other makes. None of the participants used headphones. Response data was coded as *Yanny*, *Laurel*, or ‘mixed’ (which included responses such as ‘none of the above’, ‘both’, *Yowrie*, *Yowrel*, *Yerry*, etc.). Nominal logistic regressions were run with ‘*Yanny*, *Laurel*, mixed’ as response and age, gender, device, region, musicality, hearing impairment, English use and competence as predictors. Age*musicality, gender*device, age*device, age*region, and gender*region were included as potential interaction terms.

3. Results

The predictor of region was not significant, nor were any of the interaction terms; hearing impairment, use of English, and English competence were colinear with age, so these predictors and interaction terms were removed from the final statistical models.

3.1 Overall distribution

Overall, 62.2% reported hearing *Yanny*, 26.4% *Laurel*, and 11.4% heard a mixed version, cf. Figure 5. The multinomial logistic regression model to ascertain the role of age, gender, device, and musicality on the likelihood of a *Yanny*, *Laurel*, or

mixed percept was statistically significant ($R^2=.05$, $F(10, 974)$, $p<.001^*$). Table 1 shows the main effect likelihood ratio tests.

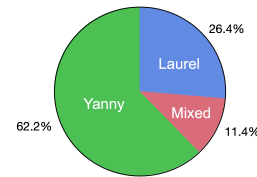


Figure 5: Distributions of *Yanny*, *Laurel*, or mixed percepts.

Predictor	DF	L-R ChiSquare	Prob>ChiSq
Age group	2	50.5295362	<.0001*
Gender	2	18.8439301	0.0001*
Device	4	18.4045911	0.0010*
Musicality	2	5.42121723	0.0665

Table 1: Main effects in the percept of *Yanny*, *Laurel* or *mixed*.

The strongest predictor of percept was age, followed by gender and device. Musicality was non-significant but just above the $p=.05$ threshold ($p=.0665$).

3.2 Age

Figure 6 shows the relative proportions of the three percepts by age group.

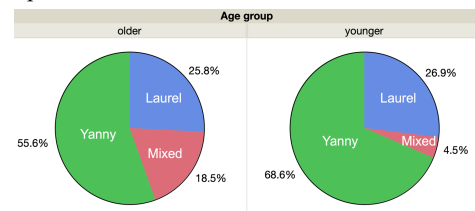


Figure 6: Distributions of percepts by age.

Older listeners (60+) had substantially more mixed percepts (18.5%) than younger participants (20-35; 4.5%) – at the cost of having much fewer *Yanny* percepts. Proportions of *Laurel* are almost identical across both age cohorts.

3.3 Gender

The results obtained on gender and percept are shown in Figure 7.

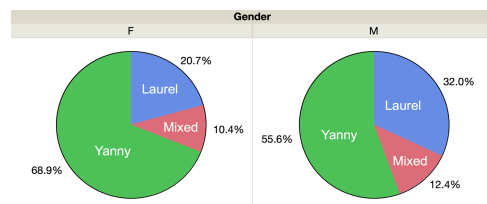


Figure 7: Distributions of percepts by gender.

While proportions of mixed percepts are relatively similar across both genders, women heard substantially more *Yanny* (68.9%) than men (55.6%).

3.4 Device

Figure 8 shows the relative proportions of *Yanny*, *Laurel*, and mixed percepts across the different devices, cf. Figure 8.

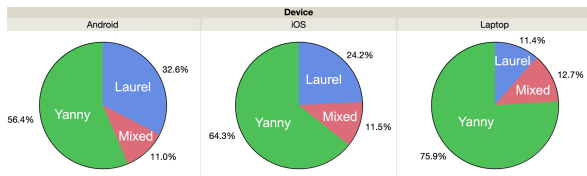


Figure 8: Distributions of percepts by device type.

What stands out in these pie charts is the varying proportions of *Yanny* and *Laurel* percepts, while the mixed category was virtually identical across all three device types. Laptop devices triggered most *Yanny* (75.9%) percepts compared to iOS (64.3%) and Android (56.4%). When looking at Android devices in a bit more detail, we find substantial variation even within this device category: 61% of the 225 Samsung device users heard *Yanny*, while only 37% of the 27 Google phone and only 29% of the 17 Xiaomi phone users heard *Yanny*.

3.5 Musicality

Although the logistic regression showed a non-significant trend toward musicality affecting percept ($p=.06$, cf. Table 1), we still feel this tendency worthy of further inspection, especially because previous literature has suggested such effects. Results revealed that musically inclined listeners heard somewhat less *Yanny* (59.7%) than non-musical listeners (65.4%).

4. Discussion

The present study was designed to determine the factors that affect the perception of *Yanny*, *Laurel*, or a mix of the two in a balanced and controlled listener sample. It was hypothesized that there would be an effect of gender, musicality, device, and region. We did find effects or tendencies for all of these factors, except for region within Switzerland. This lack of a regional effect may be indicative that segmental and phonotactic variation between dialects of Swiss German is not as vast as expected and thus do not trigger perceptual preference for phonemes or phoneme sequences familiar to the speakers [8] – unlike what has been found for typologically different languages, where native English listeners have been shown to have higher proportions of *Laurel* percepts than for example speakers of French [4]. We further expected potential effects for competence in English, which, as it turned out, was colinear with age – for which we did find an effect. In this section, we discuss these predictors in detail and offer general implications.

Overall, we found distributions of the *Yanny*, *Laurel*, and *mixed* percepts that are not consistent with previous studies. [2], for example, found a much lower proportion of *Yanny* percepts (30%), compared to 62.2% in our study. This inconsistency may have to do with differences in the phoneme inventories and phonotactics of the listeners' L1s in [2] (which are, unfortunately, not specified) and Swiss German, which may trigger higher proportions of one percept or another. Or it may have to do with the fact that [2] may have sampled more younger males (which showed a higher proportion of *Laurel* in our study). Unfortunately, [2] do not know the age and gender distributions of their participants. Generally, however, our distributions confirm that an ambiguous sound signal (which is made further ambiguous due to its noisy nature) can obscure people's perception of formants – in the same way that a phone call on a low-quality line can lead to misperceptions of sounds.

In terms of factors affecting the percept, one finding that we did anticipate – although not with such a high degree of

explanatory power – was the effect of age. The current study found older listeners to have substantially more mixed percepts at the cost of having much fewer *Yanny* percepts. Humans typically hear within the range of 20-20K Hz – that is, younger people tend to have this range. In older people, this range is narrower due to presbycusis: decaying of the hair cells in the inner ear which process higher frequencies (cf. [15, 16]). This is probably why younger listeners tend to hear more *Yanny*, as they hear the stimulus with higher fidelity in the higher frequency regions. If hearing in these regions is 'impaired', then a mixed or low-frequency percept (i.e., *Laurel*) is more likely.

As for the other factors – gender, device, and musicality – we can only provide cautious interpretations. Recall that women heard proportionately more *Yanny*, as did listeners who heard the stimulus from a laptop as well as non-musical listeners (the same gender and musicality effects were both reported by [4], although they refrained from attempting to explain the effects). Given the fact that men tend to have lower f_0 than women (cf. [17, 18]), they are perhaps more familiar with lower frequency signals (as they hear themselves talk). This 'overrepresentation' of low frequency exposure may attune their perception to certain acoustic patterns, such as the lower frequency signals as found in *Laurel* (cf. [19]); the experience one has affects the interpretation of ambiguous stimuli (cf. [20]). The device effect is contrary to our expectations, however: we expected larger speakers built into laptops – which generally transmit lower frequencies more clearly – to trigger higher proportions of *Laurel* percepts [21], which was not the case. It is difficult to explain this result; maybe it has to do with substantial between-device variation of the laptops used or perhaps some laptop fans were on during elicitation, thus obscuring sound quality (cf. [22]). Finally, musically inclined listeners tended to have proportionally more mixed and *Laurel* percepts. [23] has shown that musical and non-musical listeners exhibit differences in the processing of f_0 and spectral information. Perhaps it is these differences in processing ability that contribute to the different proportions of percepts between the two cohorts. Finally, one may argue that some of our participants may have been exposed to the stimulus (made public in May 2018) before taking part in our experiment (2020-2021). Anecdotally, however, almost none of our participants reported having heard the stimulus before.

It is currently assumed that about 2% of the vocabulary of the English language is 'polyperceivable' [24]. Stimuli such as *Yanny* v. *Laurel* are fascinating because (a) they bring speech perception and cognitive science into the limelight and (b) they trigger research such as the present study which ultimately can enhance our understanding of the mechanisms governing speech perception.

5. Acknowledgements

This research was supported by the Swiss NSF (PCEFP1_181090, dialektatlas.ch). Several people played a decisive role in collecting data to whom we are indebted (in alphabetical order): Jonathan Blum, Lara Grunder, Michelle Käch, Thea Maser, Jan Messerli, Yara Miescher, Laura Müller, Linus Oberholzer, Viviane Stebler, Melanie Studerus, Janka Szücs, Manuela Troxler, Selma Vonlanthen, Nina Von Allmen, Jessica Wagner. We further thank Thomas Kettig for editing and comments that greatly improved the manuscript.

6. References

- [1] L. Hugarass, J. Slavikova, M. Horvat, A. Al Musawi, and D. Crewther, "Temporal brightness illusion changes color perception of 'the dress'," *Journal of Vision*, vol. 17, no. 5, pp. 6, Jun. 2017.
- [2] H. R. Bosker, "Putting Laurel and Yanny in context," *The Journal of the Acoustical Society of America*, vol. 144, no. 6, pp. EL503–EL508, Dec. 2018.
- [3] G. E. Peterson, and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America*, vol. 24, no. 2, pp. 175–184, Mar. 1952.
- [4] D. Pressnitzer, J. Graves, C. Chambers, V. De Gardelle, and P. Egré, "Auditory perception: Laurel and Yanny together at last," *Current Biology*, vol. 28, no. 13, pp. R739–R741, Jul. 2018.
- [5] Slate (2018). <https://slate.com/technology/2018/05/all-the-explanations-youve-heard-so-far-about-Laurel-vs-Yanny-are-probably-wrong.htmls>
- [6] M. Kramps (2021, Oct. 10). *Laptops with the Best Speakers in 2022*. <https://www.soundgoblin.com/laptops-with-best-speakers-2021/>
- [7] SDS = Sprachatlas der deutschen Schweiz. Bern (I–VI)/Basel: Francke (VII–VIII); 1962–2003.
- [8] P. W. Jusczyk, P. A. Luce, and J. Charles-Luce, "Infants' sensitivity to phonotactic patterns in the native language," *Journal of Memory and Language*, vol. 33, no. 5, pp. 630, Oct. 1994.
- [9] M. L. G. Lecumberri, M. Cooke, and A. Cutler, "Non-native speech perception in adverse conditions: A review," *Speech communication*, vol. 52, no. 11–12, pp. 864–886, Nov. 2010.
- [10] P. Boersma, and D. Weenink, *Praat: doing phonetics by computer [Computer program]* 2022.
- [11] P. Jeszenszky, C. Steiner, and A. Leemann, "Reduction of Survey Sites in Dialectology: A New Methodology Based on Clustering," *Frontiers in Artificial Intelligence*, vol. 4, May 2021.
- [12] A. Leemann, P. Jeszenszky, C. Steiner, M. Studerus, J. Messerli, (ongoing). *SDATS Corpus – Swiss German Dialects Across Time and Space*. Retrieved from osf.io/s9z4q
- [13] A. Leemann, P. Jeszenszky, C. Steiner, M. Studerus, and J. Messerli, "Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing," *Linguistics Vanguard*, vol. 6, no. s3, Sep. 2020.
- [14] C. Draxler, and K. Jänsch, "SpeechRecorder - a Universal Platform Independent Multi-Channel Audio Recording Software," *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. May 2004.
- [15] G. A. Gates, and J. H. Mills, "Presbycusis," *Lancet*, vol. 366, no. 9491, pp. 1111–1120, Sep. 2005.
- [16] M. C. Liberman, "Noise-Induced and Age-Related Hearing Loss: New Perspectives and Potential Therapies," *F1000 Research*, vol. 6, no. 927, Jun. 2017.
- [17] D. Kovačić, and E. Balaban, "Voice Gender Perception by Cochlear Implantees," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 762–775, Aug. 2009.
- [18] D. Maurer, N. Cook, T. Landis, and C. d'Heureuse, "Are Measured Differences Between the Formants of Men, Women and Children Due to F0 Differences?" *Journal of the International Phonetic Association*, vol. 21, no. 2, pp. 66–79, Dec. 1991.
- [19] A. M. Treisman, "Strategies and Models of Selective Attention," *Psychological review*, vol. 76, no. 3, pp. 282, May 1969.
- [20] P. Wallisch, and M. Karlovich (2019, Aug. 12). *Disagreeing about Crocs and socks: Creating profoundly ambiguous color displays*. <https://doi.org/10.31234/osf.io/zpqnv>
- [21] J. Matthes (2022, Feb. 25). *Are Heavier and Bigger Speakers Better?* <https://thehometheaterdiy.com/heavier-bigger-speakers/>
- [22] Soundgoblin (2022). www.soundgoblin.com
- [23] J. Kuang, "The Effect of Musicality on Cue Selection in Pitch Perception," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3818–3818, 2017.
- [24] K. Chandra, C. Kabaghe, and G. Valiant, "Beyond Laurel/Yanny: An Autoencoder-Enabled Search for Polyperceivable Audio," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, vol. 2, pp. 593–598, Aug. 2021.