

## **Distant reading intimate encounters: a big data approach to online erotica**

Alon Lischinsky and Kat Gupta (Oxford Brookes University, UK)

Once imprisoned in “secret museums” and hidden from the view of the general public (Kendrick, 1997), pornography has become an increasingly visible and important part of cultural life over the past 50 years. Visual and written representations of sexual activity, formally banned as obscene in most Western countries since the mid-19th century, entered the mass market in the 1960s and their circulation grew very significantly with the proliferation of special-interest magazines in the 1970s, the launch of home video systems in the 1980s, and the commercial internet in the 1990s (Hardy, 2009). Online pornography in particular has disrupted the various barriers historically erected to regulate access to erotic materials, giving rise to heated discussions about acceptable forms of sexual knowledge, sexual freedom and sexual representations (Atwood, 2010).

This newfound visibility has resulted in a dramatic increase in the amount and range of scholarly work on porn. While academic research on the subject until the 1990s tended to focus on alleged undesirable effects of porn consumption —such as undermining traditional values of monogamy and emotional attachment (Zillmann, 1986), or enticing men to sexual violence against women (Mackinnon 1989)—current work adopts a much more nuanced view of the various forms in which porn is consumed, of its psychological and social functions, and of its aesthetic and cultural significance (Wicke, 1991). But while this scholarship has led to increasing awareness of the various forms of pornographic expression, it has largely focused on the visual genres of photography and film, and exploration of the language of contemporary porn remains limited and uneven (Wicke, 1991:75).

Specific forms of erotic writing have received some critical attention when they intersect with established research traditions; for example, work from a lavender linguistics perspective has investigated the ways in which gay, lesbian and queer identities are constructed in pornographic genres (Baker, 2005; Bolton, 1995; Jacobs, 2000; Koller, 2015; Morrish & Sauntson, 2007), while scholars of a primarily historical and literary bent have explored the evolving shapes of pornographic genres (Frantz, 1989; Hunt, 1993; Moulton, 2000; Viridis, 2014). But, somewhat paradoxically, research on contemporary, heterosexual, mainstream pornography has been scarce (among the few exceptions, see Marko, 2008).

Such an omission is especially unfortunate because any attempt to interpret the language used to depict sexuality and sexual activity, and especially its relation to contentious social issues, is necessarily probabilistic and comparative (Partington et al., 2013:12). Any principled attempt to evaluate contentions such as that contemporary pornographic texts are characterised by “representational practices that demean women” (Jeffries, 2007:1) requires the comparison of a corpus of such texts with references built from analogous forms of language use (such as fiction in general, or even specific genres showing specific affinities such as romance fiction; cf. Patthey-Chavez et al., 1996). In the absence of a systematic description of overall patterns in erotic writing as a whole, any claims about the relevance of the linguistic phenomena observed to occur in specific cases or subtypes are essentially

speculative, and risk being biased by the drive towards corroborating previous hypotheses (Marchi & Taylor, 2009).

Equally important is that such descriptions should avoid the temptation to reduce genres as a homogeneous monolith. Partington et al. (2013:302) have pointed out that typical pairwise comparison techniques such as keywords are intrinsically geared towards identifying inter-corpus differences, and conversely neglect similarities between corpora. In a similar manner, these techniques are of little help in detecting or measuring internal diversity within a corpus, and as a result may foster the appearance of uniformity or homogeneity within predefined categories (Gries, 2006). In the light of the ongoing diversification of contemporary pornography, characterised by 'diff'rent strokes for diff'rent folks' (Mazières et al., 2014:80), reducing the key features of erotic writing to just a central tendency would obscure the variety of sexual and stylistic interests that underlie it.

In this paper, we employ two sets of techniques for dimensionality reduction to attempt a systematic description of this diversity. On one hand, we employ topic modeling tools borrowed from computational linguistics (Jaworska & Nanda, 2016) to produce data-driven, unsupervised classifications of stories collected from Literotica.com, one of the oldest and largest erotic fiction repositories online containing approximately 1.25 billion words. On the other hand, we employ the resampling and bootstrapping techniques described by Gries (2006, 2008) to attempt to measure the degree of homogeneity in the corpus and determine an appropriate level of granularity for classification. Using co-occurrence measures, we show that the diversity of content in the stories is not easily captured by pre-set ontologies such as the genre system employed by the repository, but rather reflects a complex mixture of semantic and stylistic traits that require detailed linguistic analysis.

## References

- Attwood, F. (2010). Porn studies: from social problem to cultural practice. In F. Attwood (Ed.), *porn.com: Making Sense of Online Pornography*. Oxford: Peter Lang (pp. 1–13). New York: Peter Lang.
- Baker, P. (2006). *Public discourses of gay men*. Abingdon & New York: Routledge.
- Bolton, R. (1995). Sex talk: Bodies and behaviors in gay erotica. In: Leap, W. (ed). *Beyond the Lavender Lexicon. Authenticity, Imagination, and Appropriation in Lesbian and Gay Languages* (pp. 173-206). Amsterdam: Gordon and Breach.
- Carter, R. (1997). *Investigating English Discourse: Language, Literacy, Literature*. London: Routledge.
- Frantz, D. O. (1989). *Festum voluptatis: a study of Renaissance erotica*. Columbus, OH: Ohio State University Press.
- Gries, S. T. (2006). Exploring variability within and between corpora: some methodological considerations. *Corpora*, 1(2), 109-151.
- Gries, S. T. (2008). Corpus data in usage-based linguistics: what's the right degree of granularity for the analysis of argument structure constructions? In: Brda, M., & Žic Fuchs, M. (eds). *Expanding Cognitive Linguistic Horizons* (pp. 237–256). Amsterdam: John Benjamins.
- Halliday, M. A. K. (1973). *Explorations in the functions of language*. London: Edward Arnold.

- Hardy, S. (2009). The new pornographies: Representation or reality? In F. Attwood (Ed.), *Mainstreaming Sex: The Sexualization of Western Culture* (pp. 3–18). London: IB Taurus & Co.
- Hunt, L. A. (Ed.). (1993). *The invention of pornography: obscenity and the origins of modernity, 1500-1800*. New York: Zone
- Jacobs, G. (2000). Applying functional grammar: A discourse analysis of sexually explicit stories written in adult magazines. *Social Semiotics*, 10(3), 281-292.
- Kendrick, W. (1997). *The Secret Museum: Pornography in Modern Culture*. Berkeley, CA: University of California Press.
- Koller, V. (2015). The subversive potential of queer pornography. A systemic-functional analysis of a written online text. *Journal of Language and Sexuality*, 4(2), 254–271.
- MacKinnon, C. A. (1989). *Toward a feminist theory of the state*. Cambridge, MA: Harvard University Press.
- Marko, G. (2008). *Penetrating language: a critical discourse analysis of pornography*. Tübingen: Narr.
- Morrish, E., & Sauntson, H. (2007). *New Perspectives on Language and Sexual Identity*. Basingstoke & New York: Palgrave Macmillan.
- Moulton, I. F. (2000). *Before Pornography: erotic writing in early modern England*. Oxford & New York: Oxford University Press.
- Trew, T. (1979). "What the Papers Say": linguistic variation and ideological difference. In R. Fowler, B. Hodge, G. Kress, & T. Trew (Eds.), *Language and control* (pp. 117–156). London: Routledge.
- Viridis, D. F. (2015). Eroticizing female and male bodies: a linguistic investigation of a pornographic novel from the Victorian magazine *The Pearl*. *Porn Studies*, 2(1), 19-34.
- Wicke, J. (1991). Through a Gaze Darkly: Pornography's Academic Market. *Transition*, (54), 68–89. doi:10.2307/2934903
- Zillmann, D. (1986). *Effects of Prolonged Consumption of Pornography*. Paper written for the Surgeon General's Workshop on Pornography and Public Health. Arlington, VA. Retrieved from [http://sgreports.nlm.nih.gov/NN/B/C/K/V/\\_/nnbckv.pdf](http://sgreports.nlm.nih.gov/NN/B/C/K/V/_/nnbckv.pdf)