# RePair and All Irreducible Grammars are Upper Bounded by High-Order Empirical Entropy

Ochoa, Carlos

Navarro, Gonzalo

© 1963-2012 IEEE. Irreducible grammars are a class of context-free grammars with well-known representatives, such as Repair (with a few tweaks), Longest Match, Greedy, and Sequential. We show that a grammar-based compression method described by Kieffer and Yang (2000) is upper bounded by the high-order empirical entropy of the string when the underlying grammar is irreducible. Specifically, given a string S over an alphabet of size sigma , we prove that if the underlying grammar is irreducible, then the length of the binary code output by this grammar-based compression method is bounded by |S|H-{k}(S) + o(|S|log sigma) for any kin o(log -sigma |S|) , where H-{k}(S) is the k -order empirical entropy of S. This is the first bound encompassing the whole class of irreducible grammars in terms of the high-order empirical entropy, with coefficient 1.