ingrain frequency distribution per corpus language and media type.

