

Distribution of Non-Sentential Utterances across BNC Genres: A Preliminary Report

Kwong-Cheong Wong

Université Paris-Diderot (Paris 7)
Paris, France

wongkwongcheong@gmail.com

Jonathan Ginzburg

Université Paris-Diderot (Paris 7)
Paris, France

yonatan.ginzburg@univ-
paris-diderot.fr

1 Introduction

One of the most important features of conversation is that it is genre-based. It is the distinctive framework of a specific genre that establishes the normative expectations which underpin, and which actually drive, the interaction in a conversation. Ginzburg (2012), building on the works of others, formulates a theory of meaning for non-sentential utterances (NSUs) and gives a rudimentary treatment of genre specificity. A step forward related to incorporating fully-fledged genre specificity into that theory of meaning is to study the distribution of NSUs across different genres and to identify the distinctive characteristics of NSUs in each distinct genre. This paper is a preliminary report on an on-going empirical study concerning NSUs in the genres of the British National Corpus (BNC).

2 Method and Findings

There are totally 23 (classified) genres in the 10-million-word spoken part of BNC (see the left column of Table 1); for the rationale behind this categorization, see Lee (2002). The sizes of these genres vary greatly, ranging from 15,105 words of *S_lect_commerce* (lecture on commerce) to 4,206,058 words of *S_conv* (causal conversation). In order to cover the NSU phenomena as many as possible in those smaller genres and to make the study feasible, files of total size in the range of 15,000-19,999 words are randomly selected from each genre. The selected sub-corpus consists of 69 files, totalling 383,979 words.

Table 1 below shows the frequency distribution of NSUs across different BNC genres in terms of percentage which is calculated, for each genre, by the formula: (total number of NSUs / total number of sentence units) x 100%. As to be expected, those genres which are more interactive in nature (e.g., interview, medical consultation, classroom, causal conversation, demonstration,

and broadcast discussion) have high frequencies of NSUs (more than 20%), whereas those genres which are not interactive in nature (e.g., broadcast news, parliament, and sermon) have few NSUs.

BNC Genre	Frequency of NSUs
<i>S_interview_oral_history</i> (Hist)	33.6%
<i>S_consult</i> (Cons)	27.8%
<i>S_classroom</i> (Class)	27.2%
<i>S_interview</i> (Intv)	26.6%
<i>S_conv</i> (Conv)	25.7%
<i>S_demonstratn</i> (Demo)	22.7%
<i>S_brdcast_discussn</i> (Discn)	22.4%
<i>S_meeting</i> (Meet)	15.8%
<i>S_courtroom</i> (Court)	15.4%
<i>S_lect_humanities_arts</i> (H_arts)	14.5%
<i>S_tutorial</i> (Tut)	14.3%
<i>S_pub_debate</i> (P_deb)	13.3%
<i>S_speech_unscripted</i> (Sp_us)	13.1%
<i>S_lect_polit_law_edu</i> (P_law)	8.2%
<i>S_speech_scripted</i> (Sp_s)	6.1%
<i>S_lect_nat_science</i> (Nat_sc)	6.0%
<i>S_lect_soc_science</i> (Soc_sc)	5.1%
<i>S_brdcast_documentary</i> (Doc)	2.0%
<i>S_sportslive</i> (Sport)	1.7%
<i>S_lect_commerce</i> (Comm)	1.2%
<i>S_brdcast_news</i> (News)	0%
<i>S_parliament</i> (Prmnt)	0%
<i>S_sermon</i> (Sermn)	0%

Table 1: Distribution of NSUs across BNC Genres

It is interesting, however, to observe that there is a significant variation of NSUs among the five lecture genres, ranging from 1.2% of *S_lect_commerce* to 14.5% of *S_lect_humanities_arts*.

This study is based on the NSU taxonomy given by Fernández and Ginzburg (2002), which consists of 15 classes of NSUs covering various kinds of acknowledgments (plain acknowledgment, repeated acknowledgement), questions (clarification ellipsis, sluice, check question), answers (short answer, plain affirmative answer, repeated affirmative answer, propositional modifier, plain rejection, helpful rejection), and exten-

sions (factual modifier, bare modifier phrase, conjunction + fragment, filler). Tables 2, 3 and 4 below show the composition of NSUs in each BNC genre.

	Ack	RepAck	CE	Sluice	CheckQ
Hist	25.50%	1.80%	0.70%	0.00%	0.20%
Cons	18.90%	0.70%	1.10%	0.10%	0.60%
Class	14.90%	2.60%	0.50%	0.30%	1.00%
Intv	20.50%	0.80%	0.60%	0.00%	1.30%
Conv	10.70%	0.80%	3.60%	0.70%	0.40%
Demo	9.40%	1.10%	0.10%	0.20%	0.60%
Discn	15.30%	0.90%	0.70%	0.40%	0.20%
Meet	10.20%	0.70%	0.40%	0.10%	0.50%
Court	10.10%	0.60%	0.40%	0.00%	0.00%
H_arts	10.60%	0.50%	0.40%	0.10%	0.00%
Tut	9.20%	0.40%	0.70%	0.00%	0.00%
P_deb	8.90%	0.10%	0.30%	0.00%	0.00%
Sp_us	5.20%	1.60%	0.70%	0.00%	0.20%
P_law	0.80%	1.60%	0.60%	0.00%	0.00%
Sp_s	3.10%	0.80%	0.40%	0.00%	0.00%
Nat_sc	1.60%	0.10%	0.30%	0.00%	1.60%
Soc_sc	2.30%	0.80%	0.10%	0.10%	0.00%
Doc	0.60%	0.10%	0.40%	0.20%	0.00%
Sport	0.80%	0.30%	0.00%	0.00%	0.00%
Comm	0.70%	0.20%	0.00%	0.00%	0.00%
News	0.00%	0.00%	0.00%	0.00%	0.00%
PrImnt	0.00%	0.00%	0.00%	0.00%	0.00%
Sermn	0.00%	0.00%	0.00%	0.00%	0.00%

Table 2: Composition of NSUs in BNC Genres

	ShortAns	AffAns	RepAffAns	PropMod	Reject
Hist	0.80%	1.80%	0.40%	0.10%	0.70%
Cons	0.60%	3.00%	0.50%	0.10%	1.30%
Class	4.00%	1.70%	0.20%	0.10%	0.70%
Intv	0.10%	1.40%	0.10%	0.40%	0.30%
Conv	1.70%	4.50%	0.20%	0.20%	1.70%
Demo	5.50%	2.60%	0.40%	0.10%	1.50%
Discn	1.60%	1.40%	0.20%	0.40%	0.90%
Meet	0.60%	1.80%	0.10%	0.30%	0.40%
Court	1.10%	1.30%	0.20%	0.40%	1.00%
H_arts	0.40%	1.90%	0.00%	0.20%	0.20%
Tut	0.60%	1.70%	0.50%	0.10%	0.40%
P_deb	0.30%	2.30%	0.20%	0.30%	0.60%
Sp_us	2.80%	0.90%	0.20%	0.20%	0.50%
P_law	3.90%	0.20%	0.20%	0.00%	0.00%
Sp_s	0.20%	0.90%	0.00%	0.00%	0.30%
Nat_sc	1.40%	0.90%	0.00%	0.00%	0.00%
Soc_sc	0.40%	0.60%	0.00%	0.00%	0.30%
Doc	0.20%	0.10%	0.10%	0.00%	0.20%
Sport	0.10%	0.00%	0.10%	0.00%	0.10%
Comm	0.00%	0.20%	0.00%	0.00%	0.00%
News	0.00%	0.00%	0.00%	0.00%	0.00%
PrImnt	0.00%	0.00%	0.00%	0.00%	0.00%
Sermn	0.00%	0.00%	0.00%	0.00%	0.00%

Table 3: Composition of NSUs in BNC Genres (cont.)

	HelpReject	FactMod	BareModPh	Conj+Frag	Filler
Hist	0.20%	0.40%	0.20%	0.40%	0.50%
Cons	0.30%	0.30%	0.10%	0.00%	0.40%
Class	0.20%	0.40%	0.00%	0.00%	0.50%
Intv	0.00%	0.60%	0.00%	0.40%	0.30%
Conv	0.20%	0.80%	0.10%	0.00%	0.10%
Demo	0.00%	0.20%	0.00%	0.10%	0.90%
Discn	0.00%	0.00%	0.20%	0.10%	0.10%
Meet	0.00%	0.10%	0.00%	0.10%	0.40%
Court	0.10%	0.20%	0.00%	0.00%	0.00%
H_arts	0.20%	0.00%	0.00%	0.00%	0.00%
Tut	0.10%	0.40%	0.00%	0.00%	0.20%
P_deb	0.10%	0.20%	0.00%	0.00%	0.20%
Sp_us	0.00%	0.40%	0.20%	0.00%	0.10%
P_law	0.20%	0.80%	0.00%	0.00%	0.00%
Sp_s	0.10%	0.00%	0.00%	0.10%	0.00%
Nat_sc	0.00%	0.00%	0.00%	0.00%	0.00%
Soc_sc	0.10%	0.10%	0.00%	0.00%	0.10%
Doc	0.00%	0.10%	0.00%	0.00%	0.00%
Sport	0.10%	0.00%	0.10%	0.10%	0.10%
Comm	0.00%	0.00%	0.00%	0.00%	0.00%
News	0.00%	0.00%	0.00%	0.00%	0.00%
PrImnt	0.00%	0.00%	0.00%	0.00%	0.00%
Sermn	0.00%	0.00%	0.00%	0.00%	0.00%

Table 4: Composition of NSUs in BNC Genres (cont.)

It can be observed that, except for *S_lect_polit_law_edu* (and those genres which have no NSUs), *Ack* (plain acknowledgement) is the most frequent among all NSUs in each genre. Apart from this general observation, some distinctive characteristics of some genres can also be observed. For example, *S_classroom* has a high frequency of *ShortAns* due to the fact that students frequently give answers to the teacher's questions; *S_cons* (medical consultation) has a high frequency of *AffAns* and *Reject* due to the fact that patients frequently give yes-or-no answers to the doctor's probing; *S_conv* (causal conversation), due to its causal nature, has a high frequency of *CE* (clarification ellipsis).

3 Further Work

A more in-depth study on a few selected BNC genres will be conducted.

References

- Fernández, R., & Ginzburg, J. (2002). Non-sentential utterances: A corpus-based study. *Traitement automatique des langues*, 43(2), 13-42.
- Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*, Oxford University Press.
- Lee, D. (2002). Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language and Computers*, 42(1), 247-292.