

Columbia Games Corpus

Title: Columbia Games Corpus

Author(s): Julia Hirschberg, Agustin Gravano, Stefan Benus, Gregory Ward, Elisa Sneed German.

Language(s): English (Standard American English)

Recommended/Expected use of corpus: This corpus is well-suited for studying the following topics, inter alia: the dynamics of spontaneous dyadic conversation, especially turn-taking behavior and the usage of affirmative cue words; the intonational realization of given (old, previously-mentioned) vs. new information; acoustic/prosodic entrainment (e.g. convergence, synchrony, alignment).

Collection Procedure: The Columbia Games Corpus is a collection of 12 spontaneous task-oriented dyadic conversations elicited from native speakers of Standard American English (SAE). Thirteen subjects (six female, seven male) participated in the study in October 2004. Twelve sessions, totaling 9.8h of dialogue were recorded and subsequently transcribed and annotated for various discourse/pragmatic phenomena. (See below for more detailed information).

Data Format: Audio files are in FLAC format (lossless-compressed WAV), 16 bit, mono, 16 kHz. Text files are in UTF-8 plain text.

Corpus Description

The Columbia Games Corpus is a collection of 12 spontaneous task-oriented dyadic conversations elicited from native speakers of Standard American English (SAE). It was collected and annotated jointly by the Spoken Language Group at Columbia University and the Department of Linguistics at Northwestern University, as part of an ongoing project of prosodic variation in SAE (NSF IIS-0307905).

In each session, two subjects were paid to play a series of computer games requiring verbal communication to achieve joint goals of identifying and moving images on the screen to achieve a combined number of points. Each subject used a separate laptop computer and could not see the screen of the other subject. They sat facing each other in a soundproof booth in the Speech Lab at Columbia University, with an opaque curtain hanging between them, so that all communication was verbal. The subjects' speech was not restricted in any way, and it was emphasized at the session beginning that the game was **not** timed. Subjects were told that their goal was to accumulate as many (joint) points as possible over the entire session, since they would be paid additional money for each point they earned.

Subjects and Sessions

Thirteen subjects (six female, seven male) participated in the study, which took place in October 2004 in the Speech Lab at Columbia University. Eleven of the subjects participated in two sessions on different days, each time with a different partner. All subjects reported being native speakers of

Standard American English and having no hearing impairments. Their ages ranged from 20 to 50 years (mean: 30.0; standard deviation: 10.9), and all subjects lived in the New York City area at the time of the study. They were contacted through the classified advertisements website craigslist.org. Table 1 shows detailed information of the sessions participants.

Session Number	Speaker A		Speaker B	
01	101	M	102	M
02	103	F	104	M
03	105	F	106	M
04	107	M	108	M
05	109	F	101	M
06	108	M	109	F
07	110	F	111	F
08	102	M	105	F
09	113	M	112	F
10	111	F	103	F
11	112	F	110	F
12	106	M	107	M

Table 1: Id and gender of the participants of the twelve sessions.

We recorded twelve sessions, each containing an average of 45 minutes of dialogue, totaling roughly 9 hours of dialogue in the corpus. Of those, 70 minutes correspond to the first part of the Cards game, 207 minutes to the second part of the Cards game, and 258 minutes to the Objects game. On average, the first part of each Cards game took 1.9 minutes; the second part, 5.8 minutes; and the Objects game, 21.5 minutes.

Additionally, before the actual games, subjects played one short version of each game to become familiar with the environment. The curtain was removed during these preliminary games, so there could be visual communication between the players, and they were allowed to ask questions to the experimenter. The total duration of the preliminary games was 110 minutes. This data was not used in any experiments.

Each subject was recorded on a separate channel of a DAT recorder, at a sample rate of 48kHz with 16-bit precision, using a Crown head-mounted close-talking microphone. Each session was later down-sampled to 16k, 16 bits, saved as two separate mono WAV files, one for each player, and subsequently compressed using FLAC lossless compression.

Cards Games

Subjects were first asked to play three instances of the Cards game, where they were shown cards with one to four images on them. Images were of two sizes (small or large) and various colors, and were selected to have descriptions which would be as voiced and sonorant as possible (e.g., *yellow lion*, *blue mermaid*), to improve pitch track computations. There were two parts to each Cards game, each with different rules.

In the **first part** of the Cards game, each player's screen displayed a pile of 9 or 10 cards (Figure 1). Player A was asked to describe the top card on their pile, while Player B was asked to search through their pile to find the same card, clicking a button to indicate accomplishment. This process was repeated until all cards in Player A's deck were matched. In all cases, Player B's deck contained one additional card that had no match in Player A's deck, to prevent subjects from not describing the final card.

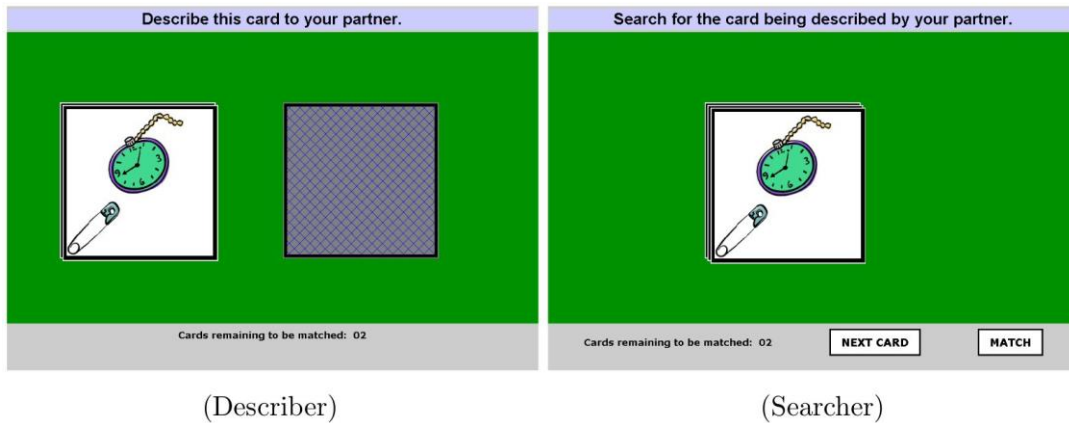


Figure 1: Sample screenshots from the first part of the Cards game.

In the **second part** of the Cards game, each player saw a board of 12 cards on the screen, all initially face down (Figure 2). As the game began, the first card on one player's (the *Describer's*) board was automatically turned face up. The *Describer* was told to describe this card to the other player (the *Searcher*), who was to find a similar card from the cards on their board. If the *Searcher* could find a card depicting one or more of the objects described by the *Describer*, the players could decide whether to declare a match and receive points proportional to the numbers of objects matched on the cards. At most three cards were visible to each player at any time, with earlier cards being automatically turned face down as the game progressed. Players switched roles after each card was described and the process continued until all cards had been described. The players were given additional opportunities to earn points, based on other characteristics of the matched cards, to make the game more interesting and to encourage discussion.

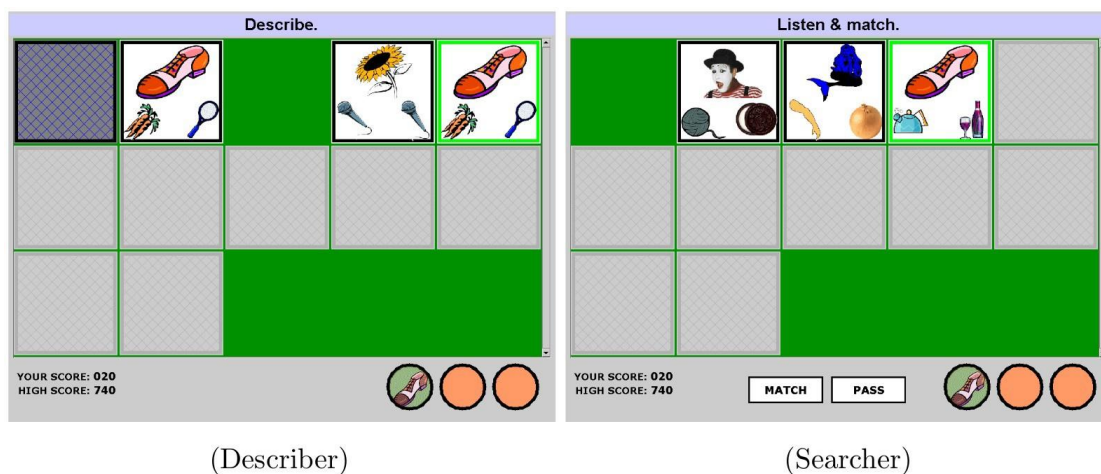


Figure 2: Sample screenshots from the second part of the Cards game.

Objects Game

After completing all three instances of the Cards game, subjects were asked to play the Objects game, which we describe in this section. As in the Cards game, all images were selected to have descriptions as voiced and sonorant as possible.

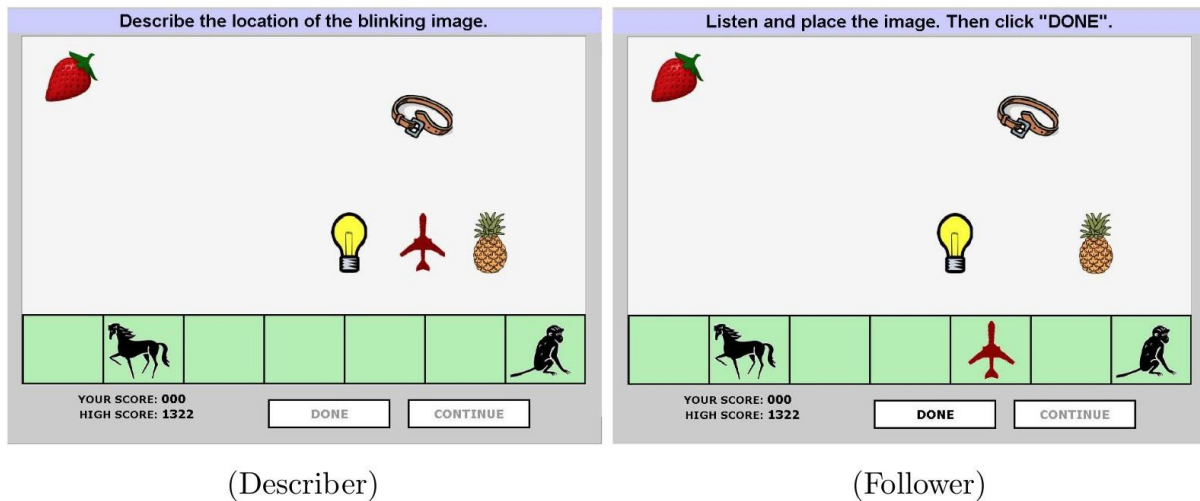


Figure 3: Sample screenshots from the Objects game.

In the Objects game, each player's laptop displayed a game board with 5 to 7 objects (Figure 3). Both players saw the same set of objects at the same position on the screen, except for one (the target). For the Describer, the target object appeared in a random location among other objects on the screen; for the Follower, the target object appeared at the bottom of the screen. The Describer was instructed to describe the position of the target object on their screen so that the Follower could move their representation to the same location on their own screen. After players negotiated their best location match, they were awarded 1 to 100 points based on how well the Follower's target location matched the Describer's.

The Objects game proceeded through 14 tasks. In the initial four tasks, one of the subjects always acted as the Describer, and the other one as the Follower. In the following four tasks they inverted their roles: the subject that played the Describer role in the initial four tasks was now the Follower, and vice versa. In the final six tasks, they alternated the roles with each new task.

Files

Each directory `session_NN/` (with NN: 01..12) contains these files:

<code>sNN.cards.1.A.flac</code>	cards game #1 - player A
<code>sNN.cards.1.B.flac</code>	cards game #1 - player B
<code>sNN.cards.2.A.flac</code>	cards game #2 - player A
<code>sNN.cards.2.B.flac</code>	cards game #2 - player B
<code>sNN.cards.3.A.flac</code>	cards game #3 - player A

```
sNN.cards.3.B.flac cards game #3 - player B
sNN.objects.1.A.flac objects games - player A
sNN.objects.1.B.flac objects games - player B
```

For each .flac file there are corresponding .words, .dm and .turns files. Additionally, there are .tasks files, which detail the structure and other information related to Cards and Objects games played by subjects during the recording sessions. The sections below describe these file types.

Orthographic Transcriptions (.words files)

Trained annotators orthographically transcribed the recordings of the entire corpus and manually aligned the words to the speech signal, yielding a total of 70,259 words and 2037 unique words.

The files with manually time-aligned orthographic transcriptions have the .words extension. Each file contains one word per line, with this format:

```
START END VALUE
```

where START and END are floats (in seconds), and VALUE is the word itself (using the special symbol # to denote a silence). These fields are separated with a single white space.

Examples:

```
0 12.1999083 #
12.1999083 12.6899083 okay
12.6899083 13.2499083 so
13.2499083 13.3399083 the
13.3399083 14.1075 mime
14.1075 14.6699083 #
14.6699083 15.280338114001939 um
```

In this example, the file starts with a silence (#) followed by the words “okay” from roughly 12.2s to 12.69s, “so” from 12.69s to 13.25s, and so on.

Discourse Markers (.dm files)

Throughout the corpus, we noted that subjects made frequent use of a particular type of discourse markers, called *affirmative cue words*: the 5456 instances of such words account for 7.8% of the total words in the corpus. The most frequent affirmative cue word in the corpus is *okay*, with 2265 instances, followed by *right* (1258), *yeah* (903), *mm-hm* (478), *alright* (236), *uh-huh* (169), *yes* (53), *yep* (47), *gotcha* (26), *yup* (11), and *huh* (10).

Since the usage of these words apparently varies significantly in meaning, we asked three labelers to independently classify all occurrences of the 11 words listed above in the entire corpus into several discourse/pragmatic functions, including acknowledgment/agreement, backchannel, and literal modifier, among others. Definitions of these functions, as well as a detailed description of the labeling task, are provided in the file `guidelines-dm.pdf`.

The files with these annotations are named `sNN.GAME.PART.PLAYER.dm`, where

```
NN = 01..12
GAME = cards | objects
PART = 1..3 if GAME=cards, or 1 if GAME=objects
PLAYER = A | B
```

Each DM label was placed approximately in the middle point of each word.

Each affirmative cue word is followed by its majority and unanimous labels. A majority label is a label chosen by at least 2 of the 3 labelers; a unanimous label is a label chosen by all 3 labelers.

Examples:

```
243.480000 266.483750 right:N:N
266.483750 269.817500 okay:A1:?
```

In the first line, a label is located at 266.48375s for word ‘*right*’, with ‘N’ as both its majority and unanimous label. At 269.8175s, the word ‘*okay*’ has ‘A1’ as its majority label and no unanimous label. Note that the first column may be safely ignored.

For further information on these labels, see:

- A. Gravano, J. Hirschberg, S. Benus, “Affirmative cue words in task-oriented dialogue”. Computational Linguistics. Vol. 38(1), pp. 1-39, March 2012.

Turn-Taking (.turns files)

We define an *inter-pausal unit* (IPU) as a maximal sequence of words surrounded by silence longer than 50 milliseconds. A *turn* is a maximal sequence of IPUs from one speaker, such that between any two adjacent IPUs there is no speech from the interlocutor. Boundaries of IPUs and turns were computed automatically from the time-aligned transcriptions. We classified the beginning of each turn in the corpus into one of several turn-taking categories, including smooth switch, overlap, interruption, butting-in, backchannel, and others. These categories are defined in the file `guidelines-turns.pdf`, along with a detailed description of the corpus annotation.

The files with turn-taking annotations are named `sNN.GAME.PART.PLAYER.turns`, where

```
NN = 01..12
GAME = cards|objects
PART = 1..3
PLAYER = A | B
```

The Objects games were labeled by two labelers (AG and EH). Each turn exchange has the label chosen by both labelers, or ‘?’ if they did not agree (which only happened in ~28 cases).

Given the high agreement obtained for the Objects Games, the Cards games were labeled by just one labeler (AG).

The files `*.misc-turns` contain comments related to the turn-taking labeling task.

Note: Some turn intervals are labeled ‘N’ or ‘L’, which are not valid turn-taking labels. Those correspond to turns occurring outside game tasks, which are ignored in most analyses.

For further information, see:

- A. Gravano, J. Hirschberg, “Turn-taking cues in task-oriented dialogue”. *Computer Speech and Language*. Vol. 25(3), pp. 601-634, July 2011.

Tasks Files - Cards Games

The files with task structure for the Cards Games, as well as other useful information, are named `sNN.cards.M.tasks`, where

`NN = 01..12`

`M = 1..3`

A ‘*task*’ is a time interval that includes all the speech from both subjects that was intended to accomplish a particular task in the game.

Intervals marked ‘*comments*’ contain any comments made by the subjects unrelated to a specific task – typically, at the beginning or end of a game part.

In intervals marked ‘*talking-to-confederate*’, one or both subjects talked to the confederate, for example to ask questions about the game rules. These intervals are typically ignored.

Line format:

`TIME0 <TAB> TIME1 <TAB> LABEL`

where

`<TAB>` Tab character (`\t`).

`TIME0,1` Start and end time of the interval.

`LABEL` One or more of the following commands, separated by semicolons (;):

`Phase1` The current task belongs to the first phase of the game, in which one subject describes the cards in their pile while the other searches for them.

`Phase2` The current task belongs to the second phase of the game, in which the two subjects play a memory card game.

`Describer:A/B` Who describes the current card.

`Card:image1,image2,image3,...` Images in the current card being described.

`Action:match/pass` [Phase2 only] Action selected by players (match two cards or pass).

`Score:points` [Phase2 only] Number of points earned in this task.

`Time-used:NUMBER` Time used by the subjects to complete this task, calculated from the moment in which they clicked 'CONTINUE' to begin the task, to the moment when they clicked 'MATCH' in Phase1, or 'MATCH' or 'PASS' in Phase2.

Matched-card:imagen1, image2, image3, . . . [Phase2 only] Images in the card
matched to the current card.

Ignore: (BEGIN, END) , (BEGIN, END) , (BEGIN, END) , . . . Any number of intervals
that should be ignored because of, e.g., subjects talking to the confederate.

Tasks Files - Objects Games

The files with task structure for the Objects Games, as well as other useful information, are named
sNN.objects.M.tasks, where

NN = 01..12

M = 1

A *task* is a time interval that includes all the speech from both subjects that was intended to
accomplish a particular task in the game.

Intervals marked *comments* contain any comments made by the subjects unrelated to a specific
task – typically, at the beginning or end of a game part.

In intervals marked *talking-to-confederate*, one or both subjects talked to the confederate, for
example to ask questions about the game rules. These intervals are typically ignored.

Format of each line:

TIME0 <TAB> TIME1 <TAB> LABEL

where

<TAB> Tab character (\t).

TIME0,1 Start and end time of the interval.

LABEL One or more of the following commands, separated by semicolons (;):

Images:imagen1, image2, image3, . . . New set of images for the following turns.

Describer:A/B Who describes the location of the target image. The other player listens
to the description, and tries to place the target image in its correct location.

Target:image Specifies the target image for the current turn.

Score:points Number of points earned in this task.

Time-used:NUMBER Time used by the subjects to complete this task, calculated from the
moment in which they clicked 'CONTINUE' to begin the task, to the moment
when they clicked 'DONE' to confirm the location of the object.

Ignore: (BEGIN, END) , (BEGIN, END) , (BEGIN, END) , . . . Any number of intervals
that should be ignored because of, e.g., subjects talking to the confederate.

Structure of the Objects Games:

1) Both players see 6 images on their screens.

1.1) Player A describes the location of the target image; Player B places it in its correct position.

1.2) Images are rearranged, and again A describes, B places.

1.3) Images are rearranged, and again A describes, B places.

1.4) Images are rearranged, and again A describes, B places.

2) Both players see 6 images on their screens.

2.1) Player B describes the location of the target image; Player A places it in its correct position.

2.2) Images are rearranged, and again B describes, A places.

2.3) Images are rearranged, and again B describes, A places.

2.4) Images are rearranged, and again B describes, A places.

3) Both players see 6 images on their screens.

3.1) Player A describes the location of the target image; Player B places it in its correct position.

3.2) Images are rearranged, and now B describes, A places.

3.3) Images are rearranged, and now A describes, B places.

3.4) Images are rearranged, and now B describes, A places.

3.5) Images are rearranged, and now A describes, B places.

3.6) Images are rearranged, and now B describes, A places.