UBA Games Corpus

Version 1.1 - April 2025

Departamento de Computación, FCEyN, Universidad de Buenos Aires
Instituto de Ciencias de la Computación, UBA/CONICET

Agustín Gravano, Juan E. Kamienkowski, Pablo Brusco
{gravano,jkamienk,pbrusco}@dc.uba.ar

The UBA Games Corpus is a collection of direction-giving monologues (both spontaneous and read), and spontaneous, task-oriented, collaborative dialogues elicited from native speakers of Argentine Spanish. It was collected and annotated by the *Laboratorio de Inteligencia Artificial Aplicada* (LIAA)[1] at the University of Buenos Aires, as part of projects PICT 2009-0026, PICT 2014-1561, UBACYT 20020090300087 and UBACYT 20020120200025BA.

1 Objects Games dialogues

The dialogues part of this corpus was inspired by the Columbia Games Corpus [3, 4, 5]. In this part, pairs of subjects were asked to play a series of Objects Games. Each subject used a separate laptop computer and could not see the screen of the other subject. Subjects sat facing each other in a booth, with an opaque curtain hanging between them, so that all communication was verbal.

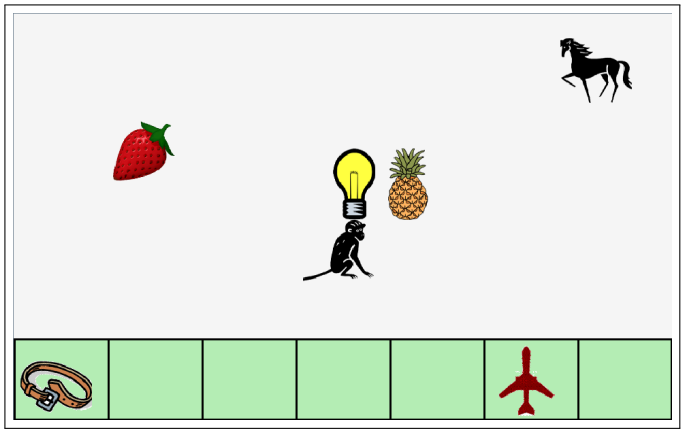

Figure 1: Sample screen from an Objects Game

In an Objects Game, each subject's laptop displayed a game board with 5–7 objects, as shown in Figure 1. Both subjects saw the same set of objects at the same position on the screen, except for one (the target). One subject (the describer) was instructed to describe the position of a target object on her screen to the other (the follower), whose task was to position the same object on her own screen. Subjects could discuss freely about the location of the target object, and were later awarded 1–100 points based

[1] http://liaa.dc.uba.ar

on how well the follower's target location matched the Describer's. Subjects were told that their goal was to accumulate as many points as possible over the entire session, since they would be paid additional money for each point they earned. Appendix A.1 includes snapshots of the instructions and sample screens of the Objects Games.

2 Direction-giving monologues

The monologues part of this corpus was inspired by the Boston Directions Corpus [7, 6]. Subjects were first asked to perform a series of five increasingly complex direction-giving tasks, ranging from explaining simple routes such as getting from one station to another on the subway, to planning a round trip journey from the University to two Buenos Aires tourist sights.

Subjects were provided with written instructions (Appendix B.1), as well as three city maps (snapshots of Google Maps) and a subway map (Appendix B.2). They were asked to address the directions to a friend of theirs, as if they were leaving a message in their friend's answering machine. Additionally, subjects were told to assume that their addressee was unaware of the descriptions given in the preceding tasks. With this procedure, we collected five spontaneous monologues from each subject.

The elicited speech was orthographically transcribed, with false starts, filled pauses and other speech errors repaired or omitted, and punctuation marks and capitalization added. All subjects returned a few days after their first session, and were asked to read their transcriptions aloud. In this way, we collected read versions of all monologues.

3 Sessions and subjects

The recordings were conducted in two batches, as summarized in Table 1. The first batch was collected in November-December, 2012, in a soundproof booth at the *Laboratorio de Investigaciones Sensoriales*[2] (Hospital de Clínicas, UBA), by Agustín Gravano. In this batch, a total of 14 subjects participated in exactly two sessions, each of which lasted approximately one hour. In their first session, each subject completed the five direction-giving tasks described above (spontaneous monologues), followed by 14 instances of the Objects Games (spontaneous dialogues), with subjects alternating in the describer and follower roles. In their second session, each subject read the transcriptions of their direction-giving tasks (read monologues), and played 14 new instances of the Objects Games (spontaneous dialogues), with a different partner. Only speech recordings were collected in this first batch.

The second batch was recorded in April, 2014, in a booth at the *Laboratorio de Neurociencia Integrativa* (Departamento de Física, FCEyN, UBA), by Juan E. Kamienkowski and Agustín Gravano. In this case, 20 subjects participated in just one session consisting of a series of 17-30 instances of the Objects Game (spontaneous dialogues). In this batch, simultaneous recordings of speech and electroencephalography (EEG) activity were collected from each participant. EEG recordings are not included in the present release, but may be shared upon request to the authors. The setup of the

[2] `http://www.lis.secyt.gov.ar`

Batch 1: 14 sessions			
5 direction-giving tasks (spontaneous)	monologue	speech	71 min
5 direction-giving tasks (read)	monologue	speech	48 min
14 instances of Objects Games	dialogue	speech	386 min
Batch 2: 10 sessions			
17-30 instances of Objects Games	dialogue	speech+EEG	320 min

Table 1: Summary of session batches

EEG equipment demanded 30 minutes. We asked subjects to complete a minimum 15 instances of the Objects Game, which typically took 15-30 minutes. After that, given the discomfort caused by the EEG equipment, we gave subjects the option to stop or continue, up to a maximum of 30 instances.

In all cases, the subjects' speech was not restricted in any way, and it was emphasized at the session beginning that the game was not timed. At the end of the session, subjects were paid a fixed amount of money for their participation, plus a bonus based on the number of points awarded in the Objects Games.

All subjects were native speakers of Argentine Spanish, lived in the Buenos Aires area at the time of the study. In the first batch of sessions (monologues and dialogues; audio only recordings), 14 subjects participated in the study (7 female, 7 male), with ages ranging from 19 to 59 years ($M = 28.6$, $SD = 12.7$); all but one subjects were right handed. In the second batch of sessions (dialogues only; audio+EEG recordings), 20 subjects took part (10 female, 10 male), age ranging between 19 and 43 years ($M = 26.4$, $SD = 6.3$), and 18 were right-handed. Detailed information about subjects may be found in the `subjects-info.csv` and `sessions-info.csv` files.

Batch 1 contains 71 minutes of spontaneous direction-giving monologue, 48 minutes of read monologue, and 386 minutes (6.4 hours) of task-oriented dialogue. Batch 2 contains 320 minutes (5.3 hours) of dialogue.

4 Corpus files

The UBA Games Corpus includes the files listed in Table 2. There is an important difference in the organization of files in batches 1 and 2. For the dialogues in batch 1, **there is one file per session (and speaker)**. In this way, the file `sNN.objects.1.{A,B}.wav` (with `NN`=01..14) corresponds to the audio for the entire session NN (and speaker A/B); this file includes all 14 game tasks in that session. The file `sNN.objects.1.{A,B}.tasks` specifies the start and end times of each game task, as explained below.

The dialogues in batch 2 are organized differently. In this case, **there is one file per game task (and speaker)**. The file `sNN.objects.TT.{channel1,channel2}.wav` (with `NN`=21..30) corresponds to the audio for task number `TT` (with `TT`=01..30), speaker A (channel1) or B (channel2). That is, the sessions are split into separate files for each game task.

Batch 1, dialogue:	
`b1-dialogue-wavs/`	Audio files.
`b1-dialogue-tasks/`	Information about the game tasks.
`b1-dialogue-words/`	Orthographic transcriptions, with temporal alignment at the word level.
`b1-dialogue-phrases/`	Orthographic transcriptions, with temporal alignment at the IPU level.
`b1-dialogue-turns/`	Turn-taking annotations.
Batch 1, monologue:	
`b1-monologue-wavs/`	Audio files.
`b1-monologue-texts/`	Orthographic transcriptions, with no temporal alignment.
Batch 2, dialogue:	
`b2-dialogue-wavs/`	Audio files.
`b2-dialogue-phrases/`	Orthographic transcriptions, with temporal alignment at the IPU level.
`b2-dialogue-turns/`	Turn-taking annotations.

Table 2: Files included in this release of the corpus.

4.1 Audio files

The audio for each subject was recorded on a separate channel of a TASCAM DR-100 digital recorder, at a sampling rate of 44.1 kHz with 16-bit precision, using a Rode HS-1 head-mounted close-talking microphone. Each channel was later downsampled to 16 kHz and saved as a separate mono wav file.

4.2 Orthographic transcriptions

All speech recordings were orthographically transcribed by trained annotators. All dialogues were manually time-aligned to the audio files at the IPU level (batches 1 and 2), and at the word level (only batch 1). An INTER-PAUSAL UNIT (IPU) is defined as a maximal speech segment from a single speaker that is surrounded by pauses longer than 100 ms.

For the monologues in batch 1, the transcriptions were made from the spontaneous direction-giving tasks. False starts, filled pauses and other speech errors were repaired or omitted, and punctuation marks were added. These transcriptions were later used for the read tasks, and have not been time-aligned to the audio files.

The time-aligned transcriptions are plain text files with one interval per line. Each interval has following format: `<start> <end> <text>`, where `<start>` and `<end>` are the time boundaries of the interval, expressed in seconds, and `<text>` is the transcription of the interval. The special symbol `#` is used for indicating a silent interval.

Example of a transcription with time-alignment at the word level:

```
b1-dialogue-words/s01.objects.1.A.words
0.000000 0.410000 #
0.410000 0.680000 bueno
```

```
0.680000 1.223913 está
1.223913 1.540000 #
1.540000 1.712226 el
1.712226 2.181653 mimo
2.181653 3.330000 #
3.330000 3.629761 está
3.629761 4.340000 titilando
4.340000 4.694193 #
4.694193 5.083623 arriba
5.083623 5.219060 de
5.219060 5.310000 la
5.310000 6.012202 lechuza
6.012202 6.330000 #
...
```

Example of a transcription with time-alignment at the IPU level:

```
b1-dialogue-phrases/s01.objects.1.A.phrases
0.0 0.41 #
0.41 1.223913 bueno está
1.223913 1.54 #
1.54 2.181653 el mimo
2.181653 3.33 #
3.33 4.34 está titilando
4.34 4.694193 #
4.694193 6.012202 arriba de la lechuza
6.012202 6.33 #
...
```

4.3 Tasks files

The files in `b1-dialogue-tasks/` (batch 1) contain the task structure for the corresponding audio file. These are plain text files with one interval per line. Each interval corresponds to a game task and has the following format: `<start> <end> <info>`, where `<start>` and `<end>` are the time boundaries of the task, expressed in seconds, and `<info>` has a sequence of one or more of the following commands, separated by semicolons:

- `Images:image1,image2,image3,...` - Image set used in this game task. There are three possible image sets:
  - eye, mirror, bluemoon, nun, ruler, lemon, yellowmoon;
  - yellowlion, ear, mime, nail, bluelion, owl, lawnmower;
  - yellowmermaid, mm, iron, onion, whale, bluemermaid, lime.
- `Describer:A/B` - Which player describes the location of the target image. The other one 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 (from 0 to 100).
- `Time-used:NUMBER` - Time used to complete this task, computed from the moment the players clicked 'CONTINUE' to begin the task, until when they clicked 'DONE' to confirm the location of the object. Note that this may differ from the length of the task interval, computed as `<end>` − `<start>`.

In this case, the special symbol `#` is used for indicating an interval with silence and/or comments made by the speakers in between tasks.

Example: `b1-dialogue-tasks/s01.objects.1.tasks`

```
0.000000 0.046000 #
0.046000 42.061000 Images:yellowlion,ear,mime,nail,bluelion,owl,
lawnmower;Describer:A;Target:mime;Score:99;Time-used:47.107
42.061000 45.468000 #
...
251.996490 296.680000 Images:eye,mirror,bluemoon,nun,ruler,lemon,
yellowmoon;Describer:B;Target:yellowmoon;Score:97;Time-used:45.669
296.680000 299.711000 #
...
```

The files in `b2-dialogue-tasks/` (batch 2) are somewhat different. They also contain one game task per line, but with this format: `<task_number> <info>`. In this case, `<task_number>` refers to the task number in this session, from 01 to 30, and `<info>` has the same information as described above.

Example: `b2-dialogue-tasks/s21.objects.tasks`

```
01 Images:yellowlion,ear,mime,nail,bluelion,owl,lawnmower;Describer:A;
Target:mime;Score:88;Time-used:73.109
...
06 Images:eye,mirror,bluemoon,nun,ruler,lemon,yellowmoon;Describer:B;
Target:yellowmoon;Score:94;Time-used:307.991
...
```

4.4 Turn-taking annotations

The files in `b1-dialogue-turns/` and `b2-dialogue-turns/` contain the annotations of turn-taking transition types. These are plain text files with one interval per line. Each interval corresponds to a conversational turn and has the following format: `<start> <end> <label>`, where `<start>` and `<end>` are the time boundaries of the turn, expressed in seconds, and `<label>` is a turn-taking label.

A TURN is defined as a maximal sequence of IPUs from one speaker, such that between any two adjacent IPUs there is no speech from the interlocutor. Turns were automatically delimited on the time-aligned orthographic annotations, and subsequently all transitions from one turn to the next were labeled by trained annotators following

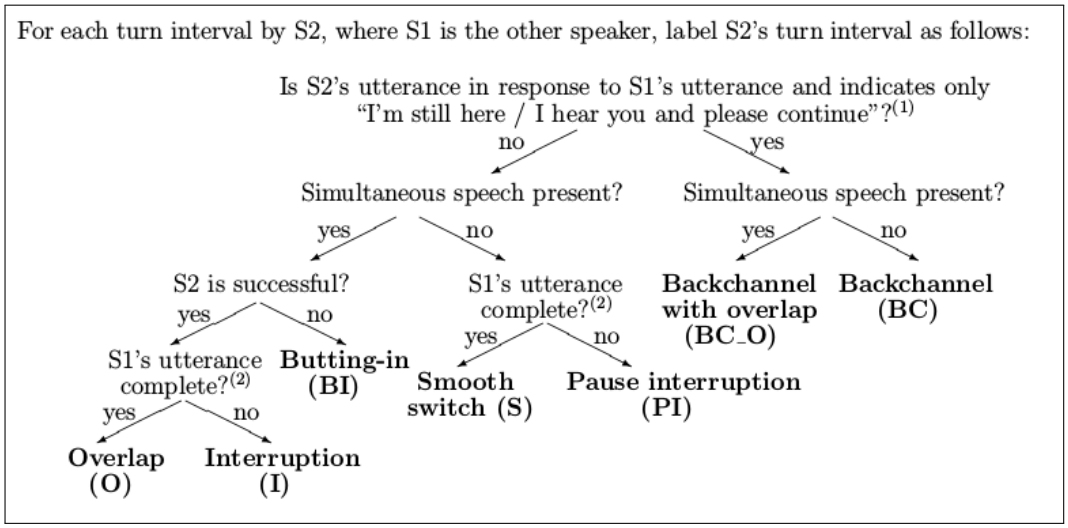

Figure 2: Turn-taking annotation guidelines

the guidelines shown in Figure 2. A more detailed description of the annotation procedure may be found in [1] and [2]. It is important to note that the HOLD transitions mentioned in these papers (i.e. when the current speaker continues talking after a short pause) are not included in these manual annotations, given that they can be trivially identified from the time-aligned transcriptions.

Example: `s01.objects.1.B.turns`

```
2.432062 2.847086 BC
2.847086 10.555597 #
10.555597 11.840000 S
11.840000 13.398680 #
13.398680 18.850000 S
18.850000 20.210000 #
20.210000 22.850361 X3
22.850361 25.036480 #
25.036480 25.744563 S
25.744563 27.264241 #
27.264241 32.590516 S
32.590516 33.474562 #
33.474562 41.652187 X3
48.226493 50.196367 X1
50.196367 51.647250 #
51.647250 52.737983 X2
52.737983 53.057480 #
...
```

In this file, each label indicates the transition type from the previous turn (by the other speaker) to the current turn. For example, the turn that starts roughly at 2.43 seconds is a backchannel, and the turn that starts at 10.56 seconds is a smooth switch.

The turns files contain a few instances of special labels: `A` represents ambiguous

transitions where annotators did not reach an agreement; `L`, `L-SIM`, `N`, and `N-SIM` are temporary labels used during the annotation procedure, left in place because they are outside game tasks (e.g., when subjects talked to the confederate between tasks). There are 402 of these special labels – 3.2% of the total turn-taking labels, and they can be safely ignored.

References

[1] Brusco, P., and Gravano, A. Automatic offline annotation of turn-taking transitions in task-oriented dialogue. *Computer Speech & Language 78* (2023), 101462.

[2] Brusco, P., Vidal, J., Beňuš, Š., and Gravano, A. A cross-linguistic analysis of the temporal dynamics of turn-taking cues using machine learning as a descriptive tool. *Speech Communication 125* (2020), 24–40.

[3] Gravano, A. *Turn-Taking and Affirmative Cue Words in Task-Oriented Dialogue.* PhD thesis, Columbia University, 2009.

[4] Gravano, A., and Hirschberg, J. Turn-taking cues in task-oriented dialogue. *Computer Speech & Language 25*, 3 (2011), 601–634.

[5] Hirschberg, J., Gravano, A., Beňuš, Š., Ward, G., and Sneed German, E. Columbia Games Corpus LDC2021S02. Web Download. *Philadelphia: Linguistic Data Consortium* (2021).

[6] Hirschberg, J., and Nakatani, C. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of ACL* (1996), pp. 286–293.

[7] Nakatani, C. H. *The computational processing of intonational prominence: A functional prosody perspective.* PhD thesis, Harvard University, 1997.

A Dialogue Materials

A.1 Instructions and screens of the Objects Games

Figure 3: Instructions of the Objects Games.

Figure 4: Describer (left) and follower screens.

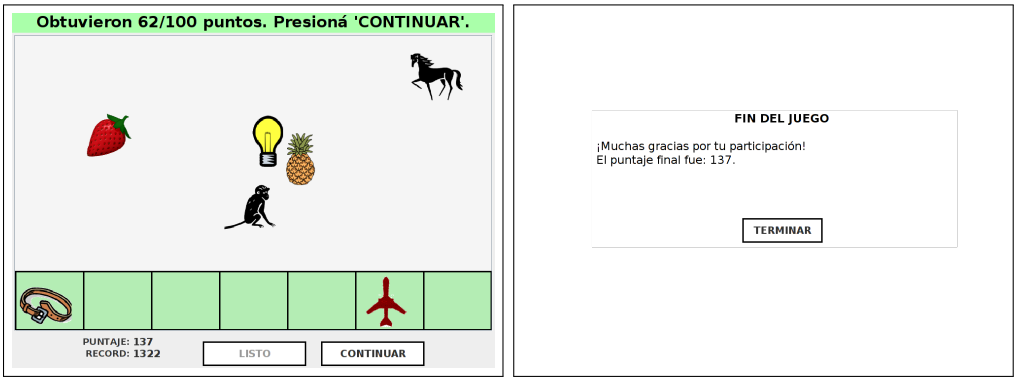

Figure 5: Screens informing the points awarded after a task, and the final score at the end of the session.

B Monologue Materials

B.1 Instructions for spontaneous monologues

These are the written instructions given to subjects for the five direction-giving tasks designed to elicit spontaneous monologues. The instructions for each task were presented separately once the preceding task had been completed.

INSTRUCCIONES

En este estudio, tu tarea consiste en explicarle a otra persona cómo llegar a determinados lugares en la ciudad de Buenos Aires, ya sea de a pie o viajando en subte. Como ayuda memoria, tendrás a tu disposición mapas del subte y de las calles de distintos puntos de la ciudad. Podés consultar los mapas cuando quieras, pero por favor no hagas ruido con las hojas mientras hablás, para no arruinar las grabaciones.

Cada tarea se describe a continuación. Leé cuidadosamente las instrucciones. Podés tomarte unos momentos para preparar el recorrido a describir. Cuando hayas terminado de preparar el recorrido, podés empezar a describirlo.

El destinatario de tus descripciones será una persona que no conoce Buenos Aires. Imaginá que le estás dejando un mensaje en el contestador de su teléfono. Además, para cada tarea, suponé que el destinatario no tiene conocimiento de tus descripciones de las tareas anteriores.

Hablá en forma natural, y sin apuro. No estamos midiendo el tiempo.

Tarea 1: Describir cómo llegar desde la estación "Facultad de Medicina" de la línea "D" del subte hasta la estación "Plaza Italia".

Tarea 2: Describir cómo llegar desde la estación "Facultad de Medicina" de la línea "D" del subte hasta la estación "Venezuela" de la línea "H".

Tarea 3: Describir cómo llegar (a pie y/o en subte) desde Callao y Santa Fe hasta el shopping del Abasto (Corrientes y Anchorena).

Tarea 4: Una amiga mendocina viene de visita para un curso en el Hospital de Clínicas. Nunca estuvo en Buenos Aires, por lo que necesita ayuda para llegar (en subte) desde la terminal de ómnibus de Retiro hasta el Hospital de Clínicas, donde la estarán esperando.

Tarea 5: Tu amiga mendocina tiene un día libre para pasear por Buenos Aires, y le gustaría conocer un poco la ciudad. Elegí 2 lugares de la siguiente lista (pueden ser los 2 lugares que mejor conozcas) y explicale a tu amiga cómo llegar a ellos. Describí un solo recorrido, que comience y termine en el Hospital de Clínicas. En lo posible, dale información adicional, como por ejemplo cómo ingresar al lugar, cosas que podría hacer allí, o detalles a los que debe prestarle atención del lugar.

- Plaza de Mayo (Cabildo, Catedral, Casa Rosada)
- Congreso Nacional
- Biblioteca Nacional
- Museo Nacional de Bellas Artes
- Museo de Arte Latinoamericano en Buenos Aires (MALBA)

- Avenida de Mayo (Café Tortoni, Palacio Barolo, 36 Billares, Teatro Avenida)
- Shopping del Abasto
- Alto Palermo Shopping
- Jardín Zoológico y/o Jardín Botánico
- Obelisco y Avenida 9 de Julio

B.2 Maps for spontaneous monologues

In addition to the written instructions, subjects were provided with three city maps and a subway map, as shown in Figures 6, 7, 8 and 9.

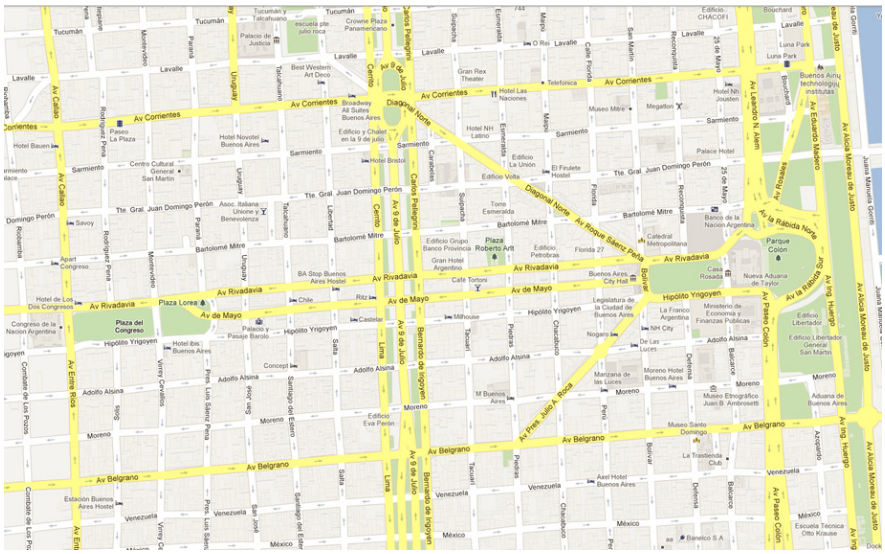

Figure 6: City map #1 given to subjects for the direction-giving tasks

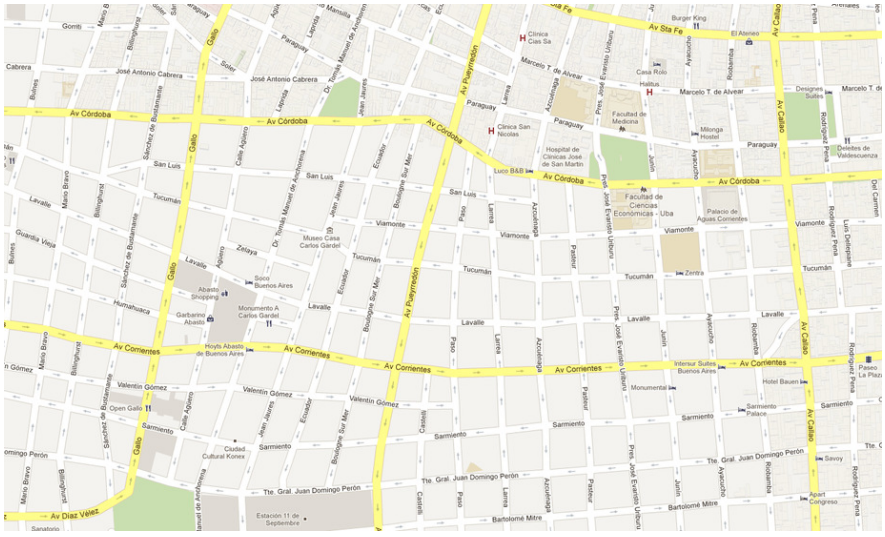

Figure 7: City map #2 given to subjects for the direction-giving tasks

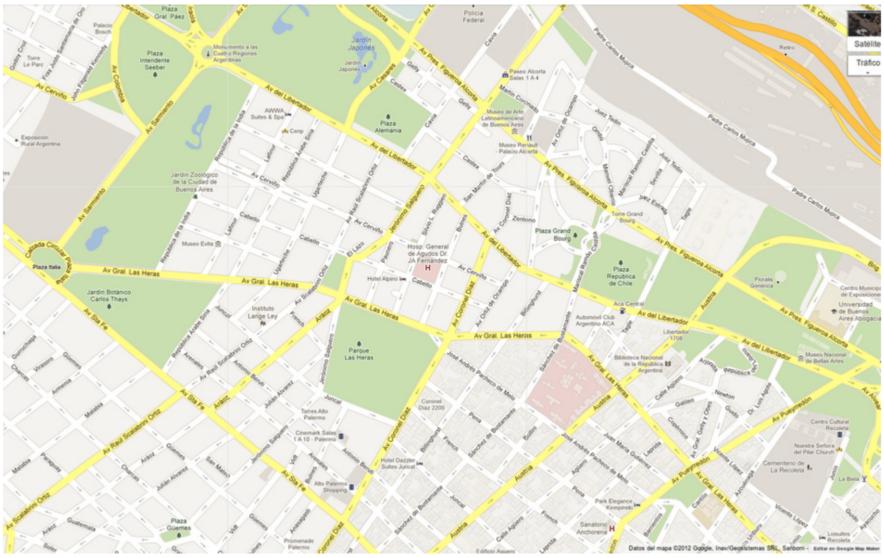

Figure 8: City map #3 given to subjects for the direction-giving tasks

Figure 9: Subway map given to subjects for the direction-giving tasks