The Slovak Autistic and Non-Autistic Child Speech Corpus: Task-Oriented Child-Adult Interactions

Joanna Kruyt*, Katarína Polónyiová†, Róbert Sabo*, Daniela Ostatníková†, Štefan Beňuš‡*

*Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
†Academic Research Center for Autism, Faculty of Medicine, Comenius University, Bratislava, Slovakia
‡ Constantine the Philosopher University in Nitra, Slovakia
{Joanna.Kruyt, Robert.Sabo}@savba.sk, {katarina.polonyiova,daniela.ostatnikova}@fmed.uniba.sk, sbenus@ukf.sk

Abstract

This paper presents the Slovak Autistic and Non-Autistic Child Speech Corpus, which consists of audio-recordings and transcripts of collaborative, task-oriented conversations between children (with or without autism spectrum disorder, ASD) and a non-autistic adult experimenter. The task used to elicit this corpus was the Maps task. This corpus was primarily recorded to investigate lexical alignment, but can also be used to study other conversation coordination strategies and behaviours. Scores on various standardised psychometric tests, such as those measuring IQ, executive functioning, and theory of mind, are included for each participant. In total, the corpus contains over 15 hours of speech. This relatively large database contains a non-Germanic language and can be shared with any qualified researcher, making it a valuable resource for replication of existing findings regarding communication and ASD as well as future research into communication between individuals with and without ASD.

Keywords: corpus, autism spectrum disorder, task-oriented

1. Introduction

Autism spectrum disorder (ASD) is associated with a range of traits, such as differences in communication. Research has for example shown that there may be differences in pragmatic processing (see e.g. Ying Sng et al. (2018) for a review) and turn-taking (e.g. Wehrle et al., 2023) between individuals with ASD and their neurotypical peers. Investigating the communication differences between people with and without ASD may provide further insights into why individuals with ASD sometimes report struggling in social settings.

One specific conversation coordination mechanism that has gained increased attention over recent years is entrainment, or the tendency of interlocutors to behave more similarly during interaction. Differences in entrainment in movement between individuals with and without ASD have been found during various tasks, for example those involving rhythmic motions such as pendulum swinging (e.g. Fitzpatrick et al., 2016) and finger tapping (e.g Koehne et al., 2016). Entrainment has also been investigated in the speech modality, and research suggests that individuals with ASD may show different strategies of entrainment on the prosodic level (e.g. Ochi et al., 2019; Lehnert-LeHouillier et al., 2020; Patel et al., 2022; see Kruyt and Beňuš (2021) for a review) and the lexical level (e.g. Stabile and Eigsti, 2022; Patel et al., 2022; Fusaroli et al., 2023). Differences in entrainment could potentially explain why mixed dyads of individuals with and without ASD may report feelings of not being on the same wavelength or "in sync".

Studies investigating lexical entrainment in individuals with ASD have employed diverse methodologies. Some studies concentrate on the entrainment of specific target words, using collaborative card-placing tasks (e.g. Slocombe et al., 2013; Branigan et al., 2016; Hopkins et al., 2017). In these tasks, experimenters use uncommon or dispreferred words to describe objects, and the extent to which individuals with ASD adopt these terms is considered a measure of entrainment. Results from such studies seem to suggest that individuals with and without ASD do not display different entrainment patterns. Importantly, conversations during these tasks are highly constrained, featuring predictable turn-taking and short exchanges. Other studies investigate entrainment during more unstructured, naturalistic conversations (e.g. Stabile and Eigsti, 2022; Patel et al., 2022; Fusaroli et al., 2023). In these studies, the proportion of shared vocabulary between participants is measured, rather than entrainment on target words. The latter group of studies suggest that individuals with ASD tend to exhibit lower degrees of lexical entrainment compared to their neurotypical peers.

In this paper, we present our Slovak Autistic and Non-Autistic Child Speech (SANACS) corpus. This was collected to combine the two approaches: en-

LREC-COLING 2024, pages 16094–16099
20-25 May, 2024.

trainment can be assessed on target words, but during a more unrestricted task. While this corpus was constructed specifically to investigate linguistic entrainment in children with and without ASD, its interactive nature allows it to be used for research into other conversation coordination strategies and mechanisms, such as for example turn-taking behaviour or backchanneling patterns.

It must be noted that recording child speech is typically more complicated than recording adults: children are more likely to not follow instructions, or to become distracted or distressed. Recording speech from children with ASD brings its own unique set of challenges: for example, differences in social interaction, sensory sensitivities, a dislike for unfamiliar environments, and an increased tendency for distress can all influence the recording process. With this in mind, we made several decisions throughout the experimental design process and recording procedure that are explained in their respective sections of this paper.

2. Related work

Several corpora of speech of individuals with ASD have been recorded, but the majority of these are not easily available. Due to more stringent privacy concerns and ethical issues, recordings from this population can often not be shared. This is especially true if the recordings are of child speech, as additional ethical constraints are placed on data from this group. Many of the available corpora dedicated to the language of individuals with ASD are listed on the TalkBank (MacWhinney, 2007) sub-section ASDBank, and include Dutch, English, French, Greek, Mandarin, and Spanish. For many of these corpora, including all English and French ones, and one in Mandarin, only transcripts are available, and access to audio or video recordings is not possible. ASDBank lists three corpora for which audio recordings are available, and one that has video as well as audio. One of the audio corpora consists of Dutch children with ASD completing a storytelling task (Hendriks et al., 2014; Kuijper et al., 2015). A corpus of audio-recordings of toy-play conversations exists for Spanish (Rodríguez-Muñoz, 2009; Muñoz, 2014). Finally, video- and audio-recordings of dialogues between a child and adult are available in Greek, though these recordings are all of the same child with ASD. In other words, most corpora are very small or do not contain naturalistic communicative interactions, and are in Germanic or Romance languages.

To sum up, many available corpora containing language from individuals with ASD consist mainly of text-based transcripts, and if audio-recordings are available, they often do not contain dialogue. The one corpus on ASDBank that contains conversations explicitly described as "dialogue" (the Greek corpus) only contains speech of one child. The available corpora are thus not ideally suited for researching conversation coordination strategies such as entrainment. Moreover, the vast majority of corpora on ASDBank or other places are from Germanic or Romance languages. The creation of a corpus in a Slavic language such as our SANACS corpus is therefore important for cross-linguistic comparisons and generalisability of theories.

Additionally, the fact that the SANACS corpus contains speech that is more naturalistic and less structured than that analysed in other studies on entrainment in individuals with ASD is an advantage: many studies rely on tasks that lead to either highly structured interaction with predictable turn-taking (e.g. Slocombe et al., 2013; Branigan et al., 2016; Hopkins et al., 2017) or use data recorded during diagnostic interviews such as the ADOS (e.g. Ochi et al., 2019), which may be experienced as stressful or emotional and might thus not reflect the way communication strategies are used in real-world, every-day conversation.

3. Methods

3.1. Corpus design and data collection

3.1.1. Speaker recruitment

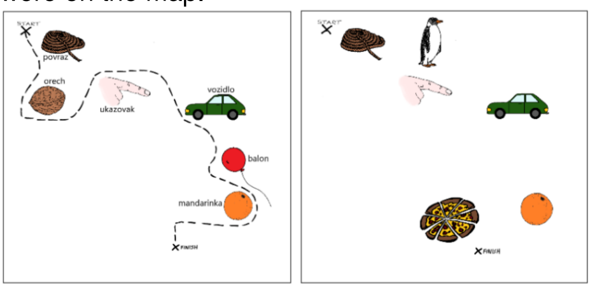
To recruit our participants and record our corpus, we collaborated closely with the Academic Research Center for Autism (ARCA) of Comenius University in Bratislava. Participants with suspected ASD were recruited from their waiting list and underwent a full diagnostic procedure to confirm their diagnosis. Neurotypical (NT) children were recruited through local primary schools and word of mouth. Note that NT children did not undergo a full ASD diagnostic procedure, but a trained psychologist performed a thorough clinical interview with the parents of each child, to rule out any developmental abnormalities as well as current psychiatric or neurological problems. Hence, there is a very small, but non-zero, chance that this group might contain undiagnosed autistic children. Our corpus was collected as part of a larger ongoing project in which various biomarkers were measured, so prepubescent children, whose hormone levels are more stable, were recruited. We aimed to match the children in our neurotypical group to our children with ASD as closely as possible in terms of gender, age, and IQ, though it is difficult to match two groups perfectly. In total, we recruited 76 participants, but data from only 67 participants can be shared: 37 children with ASD (7F, 30M) and 30 neurotypical children (7F, 23M).

3.1.2. Elicitation task

The SANACS corpus was collected primarily to study linguistic alignment in children with and without ASD during semi-naturalistic, task-oriented dialogue. To elicit conversations, we used a variation of the collaborative Maps task (Anderson et al., 1991). Both participants are given a map that portrays several landmarks. Some landmarks are on both maps, though the two maps also depict different landmarks. The two participants cannot see each other's maps. One map, given to the instructor, depicts a route. The instructor describes this route to the other player, i.e. the follower, and their joint goal is for the follower to replicate the instructor's route as closely as possible on her map.

For the collection of the SANACS corpus, we edited the original maps to keep the location of the landmarks and routes the same, but we replaced the original black-and-white landmarks with colourful pictures of objects taken from the International Picture Naming database (Szekely et al., 2004; Snodgrass and Vanderwart, 1980). All pictures were coloured in and some objects were edited slightly to be more suitable for the experiment. Some pictures, the target objects, had multiple lexical labels: a preferred name that children tend to use often and a dispreferred one that is acceptable but children are less inclined to use. The target objects were selected after an online norming study that identified objects with preferred and dispreferred terms in the population of children matching the ages of our subjects. Control objects were selected based on the fact they had one clearly preferred name and no valid dispreferred labels. The experimenter was instructed to always use the dispreferred words for the target objects, and these words were written by the objects on her maps as a reminder.

Figure 1: Example of maps used during the Maps task. On the left is the experimenter's map with words we wanted her to use written as a reminder. This example is a map where the experimenter was the instruction giver, and half of the target objects were on the map.



We selected 8 target objects and created a total of twelve maps for the task (2 sets of 6 maps). Table 1 provides information on the structure of the task

Table 1: Overview of Maps task trial structure.

Map	Instruction giver	Objects
Practice 1	Experimenter	Control only
Trial 1	Experimenter	All target objects
Trial 2	Experimenter	Half of the target objects
Practice 2	Child	Control only
Trial 3	Child	Half of the target objects
Trial 4	Child	All target objects

and the order of the trials. Note that we created two different sets so that we could counterbalance which half of the objects appeared on all maps and which half appeared only on the first and last maps.

The child and experimenter sat at opposite ends of a table, separated by an opaque curtain to prevent them from seeing each other's maps. Instructions were read aloud by the experimenter before the task. Children were given the opportunity to ask any questions or request further explanation.

In most experiments that use the Maps task, roles switch between participants on every trial. We designed the experiment such that roles switch only once, halfway through the session, rather than after every trial. Children with ASD sometimes have difficulties with task switching and we wanted to minimise any effects of possible impaired executive functioning on their performance or conversations; see the switch of roles after task M2 in Table 1. This decision might limit direct comparisons with traditional Map task studies and affect the strategies or interactions in our SANACS corpus, but we decided to prioritise the distress levels of the children.

We also decided that the experimenter always played as the instructor first to facilitate her using the dispreferred terms for objects before the child named them, and because we assumed it would be easier and less overwhelming for the children with ASD to first play as the instruction follower for a few trials. The purpose of the practice trials was to ensure that the children understood the task and their roles in the game.

3.1.3. Other measurements

Children with (suspected) ASD underwent the Autism Diagnostic Observation Schedule (ADOS, Lord et al., 2000), the gold standard for the diagnosis of ASD. Their parent(s)/caregiver(s) underwent the Autism Diagnostic Interview (Revised, ADI-R). This was done to confirm the diagnosis of ASD.

Additionally, a number of standardised tests were administered. Each participant completed a nonverbal theory of mind (ToM) test called the Comic strip task by Cornish et al. (2010), and their parent(s)/caregiver(s) were asked to fill out the Behavior Rating Inventory of Executive Function (BRIEF, Gioia et al., 2000) as a measure of their child's executive functioning. Furthermore, each participant's

Table 2: Summary of demographic information and test scores for both groups of participants.

	ASD		NT		t-test
	mean (std)	range	mean (std)	range	p
Age	9.10 (1.71)	6.14 - 12.30	9.34 (2.04)	6.18 - 12.97	>0.05
IQ	96.80 (16.81)	52 - 131	105.8 (14.84)	67 - 134	<0.05
BRIEF	67.32 (9.61)	47 - 85	58.77 (12.81)	36 - 83	<0.01
ToM	8.29 (2.14)	2 - 14	12.17 (1.74)	8 - 15	<0.001

IQ was measured using the Woodcock-Johnson test. Note that most studies use the Wechsler Intelligence Scale for Children (WISC, Wechsler, 1949) to measure IQ in children. This test is not available in Slovak, which may limit the comparability of findings based on our corpus to those of existing studies. Table 2 shows that our two groups differed significantly on all three tests.

The experimenter had also been involved in the diagnostics procedure and children were thus familiar with her. While this familiarity and the fact that the experimenter knew whether the child did or did not have ASD may influence how she interacted with them, we again opted to minimise unfamiliarity and distress for the children with ASD.

3.2. Recording procedure

We decided it was more important to have the children be as comfortable as possible than eliciting the highest quality of the acoustic signal. We conducted the recordings in a room where diagnostic assessments took place, and thus the children with ASD were familiar with it, despite its dispreferred acoustic characteristics. The room was not sufficiently insulated from outside noises (such as traffic or occasional construction) and its characteristics did not prevent reverberations.

We opted against using head-mounted microphones as some children with ASD have sensory sensitivities and may become distressed by the sensation of a headset. The primary microphones were directional table-top Rode NT-3 microphones, placed approximately 50 cm from the speaker, used for both the child and the experimenter. Additionally, we used a lavalier Sennheiser microphone for the child. This was done for two reasons: first, during the design of our microphone set-up, we tried to keep in mind that children with ASD have a tendency to vary their intensity. The clinicians we worked with reported that sometimes the children they diagnosed would drastically increase or decrease their volume over the course of a conversation. For this reason, we set the gain on the two microphones to different levels. Second, we wanted to at least partially mitigate the acoustic characteristics of the room since the lavalier microphone was expected to be less affected by reverberations and outside noise than the table-top microphones. We hoped this setup would maximise the quality and usability of the recorded acoustic data.

All three microphones were connected to a ZOOM H6 digital recording device with XLR connectors and the signal was collected at 24-bit 48kHz sampling frequency. Due to various issues with the lavalier microphone, the signal from the table-top microphones was used whenever possible despite deficiencies in clipping (due to high gain), reverberations, and outside noise. In 7 sessions, the tabletop mic was replaced by the lavalier one.

3.3. Corpus composition

3.3.1. Transcription

After converting the audio signals from both speakers to mono files, a single native Slovak speaker transcribed the Maps task trials into text using the Transcriber interface. The speech between the trials, such as discussions of the results and comparing the maps, is currently not transcribed. The Transcriber files were converted to Praat TextGrid files. The same annotator hand-corrected the temporal boundaries of inter-pausal units of speech in Praat using now the full stereo audio files from the two microphones capturing the speakers. Specifically, the annotator used the visual representation of the signal in Praat in order to improve the identification of turn-internal and inter-turn silences, and to achieve a close signal-to-text alignment of individual speakers in the vicinity of cross-talks. The annotator received guidelines and training for both text transcription and signal-to-text alignment.

3.3.2. Corpus statistics

Table 3 provides the basic description of the corpus, particularly the length of (transcribed) speech in hours, number of inter-pausal units (IPUs), number of tokens (words) excluding filled pauses and backchannels like 'mhm', and all of these separately for the experimenter (exp) and the children (child), and separately for the autistic (ASD) and neuro-typical (NT) groups.

Table 3: Corpus description; see text above.

	ASD	NT	All
Total speech (hrs)	8.81	6.37	15.18
Transcribed map trials (hrs)	7.30	5.30	12.60
Clean speech - exp (hrs)	1.78	1.24	3.02
Clean speech - child (hrs)	1.60	1.18	2.78
IPUs - experimenter	6406	4481	10887
IPUs - child	5746	4259	10005
Tokens (words) - exp	31353	23902	55255
Toens (words) - child	16747	13404	30151

4. Conclusions and future directions

The SANACS corpus allows for the comparison of communication of children with and without ASD. It must be noted that in all cases, the child is interacting with a neurotypical adult. Any differences in communication that are found between groups may be due to a mismatch in neurotypes between the adult and the child (see e.g. Milton, 2012; Sheppard et al., 2016; Heasman and Gillespie, 2018) rather than an inherent difference between NT children and those with ASD. A valuable future corpus might include dyads that are matched on neurotype, i.e. ASD-ASD and NT-NT dyads, to see whether any differences in communication originate from the neurotype-mismatch rather than an inherent "deficit" in individuals with ASD.

Nonetheless, the SANACS corpus allows for studies that may offer valuable insights into the communication between individuals with and without ASD. The fact that this relatively large database contains a non-Germanic language and can be shared with any qualified researcher makes it a valuable resource for replication of existing findings regarding communication and ASD as well as future research into communication between individuals with and without ASD.

5. Ethical considerations

In total, 76 participants were recruited. The parents or caregivers signed a general consent to record the speech of their children and analyze the speech by the experimenter's institutions for academic purposes. Subsequently the parents and caregivers were contacted again asking them to agree with distributing the recorded anonymized speech to other researchers, which resulted in obtaining 67 such consents.

6. Availability and access

The Slovak Autistic and Non-Autistic Child Speech Corpus (SANACS) containing .wav files and accompanying Praat TextGrids is available to qualified researchers who wish to use it for academic or research purposes. The corpus is available through ELRA at `https://catalogue.elra.info/en-us/repository/browse/ELRA-S0491/`.

7. Acknowledgements

We would like to thank all participants and their parent(s)/caregiver(s) for participating in this study. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 859588, from the Slovak Granting Agency grant VEGA 2/0165/21, and the Slovak Research and Development Agency grant APVV-21-0373.

8. Bibliographical References

Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. 1991. The hcrc map task corpus. *Language and speech*, 34(4):351–366.

Matthew P Black, Daniel Bone, Marian E Williams, Phillip Gorrindo, Pat Levitt, and Shrikanth Narayanan. 2011. The usc care corpus: Child-psychologist interactions of children with autism spectrum disorders. In *Twelfth Annual conference of the international speech communication association*.

Holly P Branigan, Alessia Tosi, and Karri Gillespie-Smith. 2016. Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(11):1821.

K Cornish, N Rinehart, K Gray, and P Howlin. 2010. Comic strip task. *Melbourne: Monash University Developmental Neuroscience and Genetic Disorders Laboratory and Monash University Centre for Developmental Psychiatry and Psychology*.

Paula Fitzpatrick, Jean A Frazier, David M Cochran, Teresa Mitchell, Caitlin Coleman, and et RC Schmidt. 2016. Impairments of social motor synchrony evident in autism spectrum disorder. *Frontiers in Psychology*, 7:1323.

Riccardo Fusaroli, Ethan Weed, Deborah F Berger, and Letitia Naigles. 2023. Both children with and without autism commonly align their language with that of their caregivers. *INSAR 2023*.

Gerard A Gioia, Peter K Isquith, Steven C Guy, and Lauren Kenworthy. 2000. *Behavior rating inventory of executive function: BRIEF*. Psychological Assessment Resources Odessa, FL.

Brett Heasman and Alex Gillespie. 2018. Perspective-taking is two-sided: Misunderstandings between people with asperger's syndrome and their family members. *Autism*, 22(6):740–750.

Petra Hendriks, Charlotte Koster, and John CJ Hoeks. 2014. Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, cognition and neuroscience*, 29(4):391–407.

Zoë Hopkins, Nicola Yuill, and Holly P Branigan. 2017. Inhibitory control and lexical alignment in children with an autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 58(10):1155–1165.

Svenja Koehne, Alexander Hatri, John T Cacioppo, and Isabel Dziobek. 2016. Perceived interpersonal synchrony increases empathy: Insights from autism spectrum disorder. *Cognition*, 146:8–15.

Joanna Kruyt and Štefan Beňuš. 2021. Prosodic entrainment in individuals with autism spectrum disorder. *Topics in Linguistics*, 22(2).

Sanne JM Kuijper, Catharina A Hartman, and Petra Hendriks. 2015. Who is he? children with asd and adhd take the listener into account in their production of ambiguous pronouns. *PloS one*, 10(7):e0132408.

Heike Lehnert-LeHouillier, Susana Terrazas, and Steven Sandoval. 2020. Prosodic entrainment in conversations of verbal children and teens on the autism spectrum. *Frontiers in Psychology*, 11:582221.

Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H Cook, Bennett L Leventhal, Pamela C DiLavore, Andrew Pickles, and Michael Rutter. 2000. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, 30:205–223.

Brian MacWhinney. 2007. The talkbank project. In *Creating and Digitizing Language Corpora: Volume 1: Synchronic Databases*, pages 163–180. Springer.

Damian EM Milton. 2012. On the ontological status of autism: The 'double empathy problem'. *Disability & society*, 27(6):883–887.

Francisco José Rodríguez Muñoz. 2014. *Corpus oral de hablantes con desarrollo típico y síndrome de Asperger*. Logos Verlag.

Keiko Ochi, Nobutaka Ono, Keiho Owada, Masaki Kojima, Miho Kuroda, Shigeki Sagayama, and Hidenori Yamasue. 2019. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PloS one*, 14(12):e0225377.

Shivani P Patel, Jennifer Cole, Joseph CY Lau, Gabrielle Fragnito, and Molly Losh. 2022. Verbal entrainment in autism spectrum disorder and first-degree relatives. *Scientific reports*, 12(1):11496.

Francisco José Rodríguez-Muñoz. 2009. Síndrome de asperger. materiales y aproximación pragmalingüística. *Vol. VI. Valencia: Universitat de València*.

Elizabeth Sheppard, Dhanya Pillai, Genevieve Tze-Lynn Wong, Danielle Ropar, and Peter Mitchell. 2016. How easy is it to read the minds of people with autism spectrum disorder? *Journal of autism and developmental disorders*, 46:1247–1254.

Katie E Slocombe, Ivan Alvarez, Holly P Branigan, Tjeerd Jellema, Hollie G Burnett, Anja Fischer, Yan Hei Li, Simon Garrod, and Liat Levita. 2013. Linguistic alignment in adults with and without asperger's syndrome. *Journal of autism and developmental disorders*, 43:1423–1436.

Joan G Snodgrass and Mary Vanderwart. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, 6(2):174.

Mackenzie Stabile and Inge-Marie Eigsti. 2022. Lexical alignment and communicative success in autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 65(11):4300–4305.

Anna Szekely, Thomas Jacobsen, Simona D'Amico, Antonella Devescovi, Elena Andonova, Daniel Herron, Ching Ching Lu, Thomas Pechmann, Csaba Pléh, Nicole Wicha, et al. 2004. A new online resource for psycholinguistic studies. *Journal of memory and language*, 51(2):247–250.

David Wechsler. 1949. Wechsler intelligence scale for children; manual.

Simon Wehrle, Francesco Cangemi, Alicia Janz, Kai Vogeley, and Martine Grice. 2023. Turn-timing in conversations between autistic adults: Typical short-gap transitions are preferred, but not achieved instantly. *Plos one*, 18(4):e0284029.

Cheong Ying Sng, Mark Carter, and Jennifer Stephenson. 2018. A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments*, 3:2396941518803806.