

A Examples

Attribute	Value
incidentCause	nog onbekende oorzaak <i>unknown causes</i>
victimVehicle	personenwagen <i>passenger car</i>
suspectVehicle	vrachtwagen <i>truck</i>
dateTime	omstreeks 15.45 uur <i>around 3.45 pm</i>
incidentType	frontaal met elkaar in botsing <i>collided head-on</i>

↓

Door nog onbekende oorzaak kwamen een personenwagen en een vrachtwagen omstreeks 15.45 uur frontaal met elkaar in botsing.
Due to unknown causes, a passenger car and a truck collided head-on at around 3.45 pm.

Figure A: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the Dutch Incidents domain.

Attribute	Value
positionOfPlayer	middenvelder <i>midfielder</i>
assistName	Joey Suk <i>Joey Suk</i>
assistType	verlengd <i>extended</i>
goalName	aanvoerder Ars <i>captain Ars</i>
goalType	beheerst binnen schoof <i>composedly slid it in</i>

↓

Een geblokt schot werd door middenvelder Joey Suk verlengd tot bij aanvoerder Ars die beheerst binnen schoof.
A blocked shot was extended by midfielder Joey Suk to captain Ars, who composedly slid it in.

Figure B: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the Dutch Sports domain.

Attribute	Value
companyName	Ctac <i>Ctac</i>
stockChange	daalde <i>declined</i>
stockChangePercentage	0,5 procent <i>0,5 procent</i>

↓

Ctac daalde 0,5 procent op de lokale markt.
Ctac declined 0.5 percent on the local market.

Figure C: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the Dutch Stocks domain.

Attribute	Value
windAmount	zwak tot matig <i>weak to moderate</i>
windDirection	oost tot zuidoost <i>east to southeast</i>

↓

De wind is zwak tot matig en komt uit oost tot zuidoost.
There will be a weak to moderate breeze coming from east to southeast.

Figure D: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the Dutch Weather domain.

Attribute	Value
victimNumber	several
victimStatus	more serious injuries
takenToHospital	True
hospitalName	UMC

↓

Several with more serious injuries were later transported to UMC.

Figure E: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the English Incidents domain.

Attribute	Value
winLossRecord	15 of 19
teamName	Milwaukee

↓

But Milwaukee dropped 15 of 19 to begin the regular season's final month.

Figure F: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the English Sports domain.

Attribute	Value
exchangeName	Dow
stockChange	added
stockPoints	187.86
stockChangePercentage	0.7%

↓

The Dow added 187.86 points, or 0.7%.

Figure G: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the English Stocks domain.

Attribute	Value
timePoint	Overnight
temperatureCelsius	15C

↓

Overnight temperature of 15C.

Figure H: Example of a set of attribute-value pairs (top) and corresponding text (bottom) for the English Weather domain.

B Baseline Model Parameters

Setting	Value
Adam optimizer learning rate	5e-4
Network cell type	LSTM
Embedding (+cell) size	50
Batch size	20
Encoder length (max. input attribute-value pairs)	10
Decoder length (max. output tokens)	80
Max. training epochs	20
Training instances reserved for validation	2000

TGen training parameters as reported in (Novikova et al., 2017): main-sequence-to-sequence model with attention.

Setting	Value
Adam optimizer learning rate	1e-3
Embedding (+cell) size	50
Batch size	20
Training epochs	20
Encoder length (max. input tokens)	80
Training instances reserved for validation	2000

TGen training parameters as reported in (Novikova et al., 2017): reranker.

Setting	Value
Beam size	10
Reranker misfit penalty	100

TGen decoder parameters as reported in (Novikova et al., 2017).

The parameters are the same as the TGen parameters for the E2E dataset (Novikova et al., 2017). Raw strings are used for training and generation. Validation is performed on the reserved instances after each epoch using BLEU. Early stopping is applied if the top 3 BLEU results do not change for 5 epochs.

C Labels of all data types

Subdomain	Data types
Dutch sports	assistName, assistType, chanceForName, chanceForNationality, chanceForNumber, chanceForType, coachName, defendedName, disallowedGoalName, disallowedGoalType, finalScore, formationTeam, goalName, goalScore, goalType, goalkeeperName, halfTimeScore, hasLostTeam, hasScored, hasTiedTeam, hasWonTeam, homeAway, injuredName, injuryType, matchDate, matchStreakNumber, matchStreakType, matchTime, nextMatchDate, nextMatchHomeAway, nextMatchTeam, numberOfMatchGoals, numberOfMatchesPlayed, numberOfPoints, numberOfSeasonGoals, playerAge, playerName, playerNationality, positionOfPlayer, redCardName, refereeName, stadiumPlayed, substituteName, suspendedName, tackleGiverName, tackleRecipientName, teamName, teamStandings, twiceYellowName
English sports	ERA, RBI, atBatNumber, baseNumber, baseReachedNumber, baseStolen, basesRan, batterHitsTries, batterName, batterScoreNumber, battersFacedNumber, battingAverage, battingLineupNumber, catchType, catcherName, competitionName, earnedRunsNumber, errorNumber, fielderName, fielderPosition, finalScore, gameNumber, gameTally, hasLostTeam, hasScored, hasWonTeam, hitNumber, homeAway, homeRunNumber, injuryType, inningNumber, inningScore, inningsPitched, isOut, leftOnBase, locationPlayed, managerName, matchDate, matchStreakNumber, matchStreakType, numberOfStarts, onBaseNumber, outNumber, pitchCount, pitchNumber, pitchResult, pitchResultNumber, pitchType, pitcherName, pitcherRecord, pitcherSaveRecord, pitchesTotalThrown, presidentName, retireNumber, runAverage, runNumber, scoreNumber, scoreTally, standingsGames, startsNumber, stealNumber, strikeNumber, strikeOutNumber, strikeTrajectory, strikingType, teamName, teamRecord, teamStandings, throwDirection, umpireName, umpireType, unearnedRunsNumber, walkNumber, winLossRecord, winLossType, winningPercentage
Dutch/English stocks	amountNumber, companyName, exchangeName, locationName, moneyAmount, stockChange, stockChangePercentage, stockPoints, tickerName, timePoint
Dutch/English weather	cloudAmount, cloudChange, cloudType, compassDirection, gustAmount, gustChange, gustVelocity, locationArea, maximumTemperature, minimumTemperature, precipitationAmount, snowAmount, temperatureCelsius, temperatureChange, temperatureHotCold, timePoint, weatherArea, weatherChange, weatherFrequency, weatherIntensity, weatherOccurringChance, weatherType, windAmount, windChange, windDirection, windSpeedBft, windTurning, windType
Dutch incidents	dateTime, incidentCause, incidentLocation, incidentType, suspectAddress, suspectAge, suspectAmount, suspectDescription, suspectGender, suspectStatus, suspectVehicle, victimAddress, victimAge, victimAmount, victimDescription, victimGender, victimName, victimStatus, victimVehicle
English incidents	accidentAddress, accidentDate, hospitalName, numberOfRoundsFired, personnelArrivedTime, prisonName, shootingNumber, shootingType, suspectAge, suspectAgeGroup, suspectBased, suspectDescription, suspectGender, suspectHeight, suspectName, suspectNumber, suspectOccupation, suspectRace, suspectStatus, suspectVehicle, suspectWeapon, suspectWeight, takenToHospital, victimAge, victimAgeGroup, victimBased, victimGender, victimName, victimNumber, victimOccupation, victimRace, victimStatus, victimVehicle

Labels of data types used in the CACAPO dataset per subdomain.