ECON-IE Dataset Annotation Guidelines

1 Introduction

We are releasing these guidelines as a companion text to our ECON-IE dataset, released alongside EconBERTa: Towards Robust Extraction of Named Entities in Economics. They describe the thinking and process that went into creating the dataset.

Our goal is to extract information representing the causal knowledge found in impact evaluation studies in economics. To that end we'll be identifying five entity types in abstracts from economics research papers: *intervention*, *outcome*, *population*, *effect size* and *coreference*. The first four entities represent key information relevant to impact evaluation practitioners, and we add *coreference* as spans referring to a previously mentioned intervention or outcome.

2 General rules

Span annotation comprises of (1) selecting a span of tokens (determining boundaries of an entity) and (2) tagging it with an entity label. The following rules apply to all entities regardless of their type.

- Use only the context of a sentence for determining entities. The only exception is the coreference, which can refer to interventions or outcomes in other sentences.

- An entity can span over one or more consecutive tokens (words).

- An entity cannot span across sentence boundaries.

- Entity spans cannot overlap.

- Try not to include determiners (the, a), or adjective pronouns (this, its, these, such) in the span.

- Label distributive conjunctions and enumerations as a single span, otherwise as separated spans (e.g. "vitamins A, B, C and D" as a single span but "children and parents" as two).

- Annotate all mentions of entities, including in sentences providing background, conclusions, summaries, or with claims of external validity.

3 Intervention

Intervention describes a deliberate involvement in a process or system intended to influence events and/or consequences. In our context, intervention refers to the activities of a project, program, policy, or instrument in the field of development that aims to bring about change in an outcome thereby improving the conditions of a target population. Examples: *conditional cash transfer, leadership programs, household rainwater harvesting, research funding, access to microcredit.*

- Annotate program names and acronyms of interventions as interventions (and not as coreferences), e.g. *Start-and-Improve Your Business (SIYB) Program, No Child Left Behind, Progresa*

- Include generic terms in the intervention span when they describe the entity, e.g. *program, courses, lectures.* For example, in the following sentence, the generic term *courses* should also be part of the intervention span (*business training courses*):

    > *We used a randomized experiment to measure the impact of business training courses.*

    The same applies to other generic terms next to intervention descriptions such as *program* or *intervention.*

- Do not include delivery details or attributes of the interventions. For example, in the following sentence the intervention span should be *Carrying of firearms was banned* excluding the details on which day the ban took place.

    > *Carrying of firearms was banned on weekends after paydays, on holidays and on election days.*

- Annotate as interventions mentions of a treatment group by a program description, acronym, or program name; but do not include the term *group* in the span.

- Do not label the standard treatment that the control group receives as an intervention.

4 Outcome

An outcome is the likely or achieved change and effects of an intervention. Examples: *adoption of recommended animal breeding practices, institutional trust, math test scores, air pollution, wage levels, HIV treatment success, knowledge of water contamination*

- Include in the span generic references such as the word *effects* or *outcomes* when they are next to an outcome. For example:

> *The outcome we consider is* consumer behavior. Consumer behavior effects *are measured by a household survey.*

- Include words that make the outcome a quantity (e.g. *incidence of HIV* and not *HIV*) whenever possible.

- Very important: mentions of effect sizes should not be annotated as outcomes.

- Do not include population in the outcome, e.g. in *teenage fertility, child mortality* tag *teenage* and *child* as population, *fertility* and *mortality* as outcome.

5 Population

The population, sometimes referred to as population target, is the group of people on which the intervention is implemented or in which the outcome is measured. Examples: *women and girls, farmers, civil servants, teachers, women in rural areas*

- Include mentions of subgroups of the population where the intervention was applied or on whom an outcome was measured.

- If the text mentions a control group that had a different population and did not receive an intervention, do not annotate that. We are only interested in the populations that received an intervention.

- Do not annotate standalone generic references such as the word *population* with the population tag.

- Do not label standalone geographical names such as countries, regions or cities. For example, low-income households *in Nairobi.* Label geographical names only when part of greater entity or when otherwise the population wouldn't have a noun. For example, malnourished Indian children *and* rural Bangladesh

- Do not label standalone unit of intervention, e.g. *individuals, households, villages.* Label them when part of a greater entity, e.g. *individuals aged 70 years or more, remittance-receiving households.*

- Do not label treated group as population, e.g. *impregnated chaddar users* (*impregnated chaddar* is the intervention; *user* is a generic term and not labeled).

- Words defining the status of the population should be part of the population span, e.g. in *child with acute diarrhea* and *mothers suffering from anemia* the whole spans are population.

6   Effect size

Effect sizes are the mentions of quantitative measures of the magnitude of the intervention's effect in an outcome. Examples: *28 percentage points, 4.7%, 0.37 births, 102 μ/m3*

- Whenever possible, the effect size span should include not only the numeric measure but also the unit of measure, as long as this is not part of an outcome span. For example,

  > *PM2.5 exposure means were 266 and 102 μ/m3 during the trial period in the control and intervention groups, respectively.*

- Tag all the mentions of effect sizes, including mentions for different treatment groups, subgroups, control groups, or about the difference between treatment and control groups.

- Do not include in the span words indicating the direction of the effect size (example: *increment, decrease, increase*).

  > *The results indicate that primary school completion reduces teenage fertility by 0.37 births and the incidence of teenage childbearing by around 28 percentage points.*

- Statistical measures such as t-stats, p-values, or confidence intervals should not be considered effect sizes.

- Odds ratios and incidence rate ratios should be considered effect sizes.

7   Coreference

A coreference accounts for all generic expressions that relate to the same intervention or outcome entity. Examples: *intervention, project, program, outcome, effect, results, it, they*

- Tag only coreferences of interventions and outcomes.

- Mentions of an intervention by its name (e.g. *Progresa, Sesame Street*) should not be tagged as coreferences but as interventions.

- Words *treatment, intervention, experiment* in *treatment group, intervention group, experiment group* should also be annotated as coreferences excluding the word *group* in the span.

8   General rules

Span annotation comprises of (1) selecting a span of tokens (determining boundaries of an entity) and (2) tagging it with an entity label. The following rules apply to all entities regardless of their type.

- Use only the context of a sentence for determining entities. Only exception is the coreference, which can refer to interventions or outcomes in other sentences.

- An entity can span over one or more consecutive tokens (words).

- An entity cannot span across sentence boundaries.

- Entity spans cannot overlap.

- Try not to include determiners (the, a), or adjective pronouns (this, its, these, such) in the span.

- Label distributive conjunctions and enumerations as a single span, otherwise as separated spans (e.g. "vitamins A, B, C and D" as a single span but "children and parents" as two).

- Annotate all mentions of entities, including in sentences providing background, conclusions, summaries, or with claims of external validity.

9   Intervention

Intervention describes a deliberate involvement in a process or system intended to influence events and/or consequences. In our context, intervention refers to the activities of a project, program, policy, or instrument in the field of development that aims to bring about change in an outcome thereby improving the conditions of a target population. Examples: *conditional cash transfer, leadership programs, household rainwater harvesting, research funding, access to microcredit.*

- Annotate program names and acronyms of interventions as interventions (and not as coreferences), e.g. *Start-and-Improve Your Business (SIYB) Program, No Child Left Behind, Progresa*

- Include generic terms in the intervention span when they describe the entity, e.g. *program, courses, lectures.* For example in the sentence *We used a randomized experiment to measure the impact of business training courses* the generic term *courses* should also be part of the intervention span (*business training courses*). The same applies to other generic terms next to intervention descriptions such as *program* or *intervention.*

5

- Do not include delivery details or attributes of the interventions. For example, in the sentence *Carrying of firearms was banned on weekends after paydays, on holidays and on election days.* the intervention span should be *Carrying of firearms was banned* excluding the details on which day the ban took place.

- Annotate as interventions mentions of a treatment group by a program description, acronym, or program name; but do not include the term *group* in the span.

- Do not label the standard treatment that the control group receives as an intervention.

10 Outcome

An outcome is the likely or achieved change and effects of an intervention. Examples: *adoption of recommended animal breeding practices, institutional trust, math test scores, air pollution, wage levels, HIV treatment success, knowledge of water contamination*

- Include in the span generic references such as the word *effects* or *outcomes* when they are next to an outcome. For example: *The outcome we consider is **consumer behavior**. **Consumer behavior effects** are measured by a household survey.*

- Include words that make the outcome a quantity (e.g. *incidence of HIV* and not *HIV*) whenever possible.

- Very important: mentions of effect sizes should not be annotated as outcomes.

- Do not include population in the outcome, e.g. in *teenage fertility, child mortality* tag *teenage* and *child* as population, *fertility* and *mortality* as outcome.

11 Population

The population, sometimes referred to as population target, is the group of people on which the intervention is implemented or in which the outcome is measured. Examples: *women and girls, farmers, civil servants, teachers, women in rural areas*

- Include mentions of subgroups of the population where the intervention was applied or on whom an outcome was measured.

- If the text mentions a control group that had a different population and did not receive an intervention, do not annotate that. We are only interested in the populations that received an intervention.

- Do not annotate standalone generic references such as the word *population* with the population tag.

- Do not label standalone geographical names such as countries, regions or cities. Label them when part of greater entity or when otherwise the population wouldn't have a noun. E.g. label *malnourished Indian children* and *rural Bangladesh*, but in *low-income households in Nairobi* label only *low-income households*.

- Do not label standalone unit of intervention, e.g. *individuals, households, villages*. Label them when part of a greater entity, e.g. *individuals aged 70 years or more, remittance-receiving households*.

- Do not label treated group as population, e.g. *impregnated chaddar users* (*impregnated chaddar* is the intervention; *user* is a generic term and not labeled).

- Words defining the status of the population should be part of the population span, e.g. in *child with acute diarrhea* and *mothers suffering from anemia* the whole spans are population.

12   Effect size

Effect sizes are the mentions of quantitative measures of the magnitude of the intervention's effect in an outcome. Examples: *The results indicate that primary school completion reduces teenage fertility by **0.37 births** and the incidence of teenage childbearing by around **28 percentage points**. Results: PM2.5 exposure means were **266** and **102 μ/m3** during the trial period in the control and intervention groups, respectively.*

- Whenever possible, the effect size span should include not only the numeric measure but also the unit of measure, as long as this is not part of an outcome span.

- Tag all the mentions of effect sizes, including mentions for different treatment groups, subgroups, control groups, or about the difference between treatment and control groups.

- Do not include in the span words indicating the direction of the effect size (example: *increment, decrease, increase*).

- Statistical measures such as t-stats, p-values, or confidence intervals should not be considered effect sizes.

- Odds ratios and incidence rate ratios should be considered effect sizes.

13 Coreference

A coreference accounts for all generic expressions that relate to the same intervention or outcome entity. Examples: *intervention, project, program, outcome, effect, results, it, they*

- Tag only coreferences of interventions and outcomes.

- Mentions of an intervention by its name (e.g. *Progresa, Sesame Street*) should not be tagged as coreferences but as interventions.

- Words *treatment, intervention, experiment* in *treatment group, intervention group, experiment group* should also be annotated as intervention coreferences excluding the word *group* in the span.