Linguistic word game based on WordNet

Implementation of a game for L2 learners of English

Deborah Jakobi/Laura Stahlhut/ Yining Wang

Table of Contents

- Project overview
- Question creation
    - Question types
    - Manual question creation
    - Automatic question creation
- Game implementation
- Discussion
- Conclusion

Project overview

Idea & Goal

- Make use of a linguistic database (LDB) to teach people a language

- Create a game based on specific LDB (WordNet)

- Create (part of) the game automatically

- Core concerns:
  - What kinds of questions can we come up with using a linguistic database?
  - How can create the questions and wrong answers automatically?
  - What is the quality of our automatically created questions?

WordNet

- Underlying database of our game

- 117'597 Synsets (81'426 Nouns)

- Knowledge in WordNet that is important for us:
  - Word definitions
  - Word relations

Question Creation

Game idea

What kind of questions can we come up with for **Noun synsets** in WordNet?

Topics:

- Definitions of Noun synsets
- Relations between Noun synsets
  - Synonymy
  - Antonymy
  - Hyponymy
  - Hyperonymy
  - Meronymy
  - Holonymy

Question types:

- Multiple choice
- Fill-in-the-blanks
- True/False Questions

Question types

1. Multiple Choice Questions (noun definitions)

    *Q: What is the correct definition of: **table**?*
    a) a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs
    b) a seat for one person, with a support for the back (chair)
    c) a small room (or recess) or cabinet used for storage space(closet)
    d) a piece of furniture that provides a place to sleep(bed)
    → correct answer: a)

2. Fill-in-the-blank Questions (noun relations)

    *The word **'material'** is a(n) _____ of **'paper' (**a material made of cellulose pulp derived mainly from wood or rags or certain grasses)".*
    a) hyponymy
    b) hypernym
    c) meronymy
    → correct answer: b)

3. True-False Questions (noun relations)

    *"**housing**", "**lodging**", "**living accommodations**" are hypernyms of **apartment** (a suite of rooms usually on one floor of an apartment house).*
    *   *a)   True*
    *   *b)   False*
    → correct answer: a)

Manual creation of questions

Manual question creation: multiple choice questions

For English language learner: 🌟identify the meaning of words

1-3: some of the most common objects in life

example:

What is the correct definition of a: table?

a) a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs

b) a seat for one person, with a support for the back (chair)

c) a small room (or recess) or cabinet used for storage space(closet)

d) a piece of furniture that provides a place to sleep(bed)

Manual question creation: multiple choice questions (definitions)

4-9: some words that are easily confused when learning English vocabulary

example:

What is the correct definition of : aboard?

a) to or in a foreign country(abroad)

b) on a ship, train, plane or other vehicle

c) get on board of (trains, buses, ships, aircraft, etc.)(board)

d) having great (or a certain) extent from one side to the other (broad)

10: distinguish different rhetorical devices

example:

What is the correct definition of : metaphor?

a) a figure of speech that expresses a resemblance between things of different kinds (usually formed with `like' or `as')(simile)

b)substituting the name of an attribute or feature for the name of the thing itself (metonymy)

c) a figure of speech in which an expression is used to refer to something that it does not literally denote in order to suggest a similarity

d)a similar kind (like)

Manual question creation: fill-in-the-blanks questions (relations)

3 Types of relations for noun synsets in WordNet:

- hyponymy (the semantic relation of being subordinate or belonging to a lower rank or class),
- hypernymy (the semantic relation of being superordinate or belonging to a higher rank or class)
- meronymy (the semantic relation that holds between a part and the whole)

example:

The word 'dormitory' is a(n) _____ of 'building' (a college or university building containing living quarters).

- *Correct : Hyponym*
- *Wrong : Hypernym*
- *Wrong : Meronym*

B2.2: The word 'annex' is a(n) _____ of 'building' (an addition that extends a main building).

- *Wrong : Hyponym*
- *Wrong : Hypernym*
- *Correct : Meronym*

B2.1: The word 'construction' is a(n) _____ of 'building' (a thing constructed; a complex entity constructed of many parts).

- *Wrong : Hyponym*
- *Correct : Hypernym*
- *Wrong : Meronym*

- S: (n) **building**, edifice (a structure that has a roof and walls and stands more or less permanently in one place) *"there was a three-story building on the corner"; "it was an imposing edifice"*
  - *direct hyponym* / *full hyponym*
  - *part meronym*
  - *has instance*
  - *direct hypernym* / *inherited hypernym* / *sister term*

Manual question creation: true-false questions (relations)

4 Types of relations for word synsets in WordNet:

- hyponymy (the semantic relation of being subordinate or belonging to a lower rank or class),
- hypernymy (the semantic relation of being superordinate or belonging to a higher rank or class)
- meronymy (the semantic relation that holds between a part and the whole)
- antonymy(the semantic relation that holds between two words that can (in a given context) express opposite meanings)

example:

"**Felicitous**" (marked by good fortune) is the antonym of "**unfortunate**" (not favored by fortune; marked or accompanied by or resulting in ill fortune).

True

> • S: (adj) **felicitous**, happy (marked by good fortune) *"a felicitous life"; "a happy outcome"*
> - *similar to*
> - *derivationally related form*
> - *antonym*
>   - • W: (adj) unfortunate [Indirect via fortunate] (not favored by fortune; marked or accompanied by or resulting in ill fortune) *"an unfortunate turn of events"; "an unfortunate decision"; "unfortunate investments"; "an unfortunate night for all concerned"*

Automatic creation of questions

Automatic question creation: multiple choice questions (noun definitions)

1. Get word to define + correct definition

    a. get random word from Brown corpus

    b. search for appropriate synset from WordNet

2. Get **wrong** definitions

    a. get random definitions → the probability of picking a word that is **not similar** to the correct word is very high

3. Shuffle answers + keep index of correct definition

Examples

What is the correct definition of: **pressure** ?

0 : the members of a business organization that owns or operates one or more establishments
1 : **the somatic sensation that results from applying force to an area of skin**
2 : a fabric made of compressed matted animal fibers
3 : a particular seat in an orchestra

What is the correct definition of: **thing** ?

0 : a full supply
1 : a river in southern Scotland that flows eastward to the Firth of Forth
2 : United States inventor who incorporated Polaroid film into lenses and invented the one step photographic process (1909-1991)
3 : **a separate and self-contained entity**

Automatic question creation: fill-in-the-blanks (noun relations)

6 Types of relations for noun synsets in WordNet:

- synonym
- antonym
- hyponyms
- hypernym
- meronym
- holonym

- S: (n) **house** (a dwelling that serves as living quarters for one or more families) *"he has a house on Cape Cod"; "she felt she had to get out of the house"*
  - *direct hyponym* | *full hyponym*
  - *part meronym*
  - **direct hypernym** | *inherited hypernym* | *sister term*
    - S: (n) dwelling, home, domicile, abode, habitation, dwelling house (housing that someone is living in) *"he built a modest dwelling near the pond"; "they raise money to provide homes for the homeless"*
    - S: (n) building, edifice (a structure that has a roof and walls and stands more or less permanently in one place) *"there was a three-story building on the corner"; "it was an imposing edifice"*
  - *derivationally related form*

Question Format:

*"The word **Y** is a(n) _____ of **X** (definition of X)".*

- *Correct answer:* the correct relation
- *Wrong answers:* randomly generated from the remaining 5 relation types

Example Question:

*"The word **dwelling** is a(n) _____ of **house** (a dwelling that serves as living Quarters for one or more families)".*

- *Correct answer:* hypernym
- *Wrong answers:* synonym, hyponym, meronym

Automatic question creation: true-false questions (noun relations)

Question Format:

"Determine whether the following sentence is true or false: The word *Y* is a *relation a* of the word *X.*"

- X = a randomly scraped synset

- Y = a synset with the relation a to X

- relation a = a random relation from the list of 6 relations

- True if relation a = correct relation, else False

Example:

"Determine whether the following sentence is true or false: The word *stump* is a *meronym* of the word *tree.*"

Correct answer: *True*

- S: (n) **tree** (a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms)
  - *direct hyponym* / *full hyponym*
  - *part meronym*
    - S: (n) stump, tree stump (the base part of a tree that remains standing after the tree has been felled)
    - S: (n) crown, treetop (the upper branches and leaves of a tree or other plant)
    - S: (n) limb, tree branch (any of the main branches arising from the trunk or a bough of a tree)
    - S: (n) trunk, tree trunk, bole (the main stem of a tree; usually covered with bark; the bole is usually the part that is commercially useful for lumber)
    - S: (n) burl (a large rounded outgrowth on the trunk or branch of a

Game implementation

- simple GUI using pySimpleGUI

- Game flow:

  1. User chooses the settings for the game

  2. Game gets prepared, words and definitions are scraped, etc.

  3. Questions are presented to the user, sorted by type

Demo

Discussion

Comparing manually & automatically created questions

automatically created questions:

+:   create questions efficiently

-:   big difference between right and wrong answers
      Difficult to control the difficulty of the questions

manually created questions:

+:   easy to control the difficulty and domain of the questions

-:   very time-consuming

Question Types

- limited by the structure & content of WordNet

- Question types we implemented: multiple-choice, fill-in-the-blank & true-false for noun definitions and relations

- Other possibilities:
  - extend the 3 question types to both definitions and relations
  - start at a noun synset → click your way through to another synset
  - Open questions, e.g. guess a lemma from the definition

Quality of automatically created questions & (wrong) answers

Type 1: Multiple choice, definitions
- Wrong answers are often too different from the correct answer → easy choice
- Wrong answers can accidentally be a a definition that would fit
- Level of difficulty changes depending on the search term

Example:

A1.1: What is the correct definition of: **freedom** *?*

0 : the condition of being free; the power to act or speak or think without externally imposed restraints

1 : the ability and desire to purchase goods and services

2 : a militant reformer

3 : (law) a judicial division of a state or the United States (so-called because originally judges traveled and held court in different locations); one of the twelve groups of states in the United States that is covered by a particular circuit court of appeals

→ Correct answer: 0

Quality of automatically created questions & (wrong) answers

Types 2&3: Fill-in-the-blanks/True-false, relations
- Wrong answers are more suitable (easy to generate)
- Level of difficulty depends on the search term
- Not suitable for language learning, but for testing ones knowledge on linguistic concepts.

Example (Type 2):

A2.1: The word 'academic_freedom' is a(n) _____ of 'freedom' (the condition of being free; the power to act or speak or think without externally imposed restraints).
- *Antonym*
- *Hypernym*
- *Hyponym*

→ *correct: Hyponym*

Conclusion

Difficulties, insights and outlook

- Many decisions are based on randomization (automatic part)→ does not allow to control the knowledge imparted

- the definitions of some words may sometimes be too difficult for second language speakers

- Possible ways to improve and extend the game:
  - filter the language database to select target words suitable for L2 learners (e.g word lists for a certain language level)
  - Find ways to improve the wrong answers for question type 1 (e.g. similarity score)
  - extend the game to Verbs, Adjectives and Adverbs
  - Gamification (scoring, ranking)

Try it out yourself :)

https://gitlab.uzh.ch/deborah.jakobi/linguistic-databases-project

How to run the code

Set up a virtual environment

```
$ python3 -m venv wngame-env
$ source wngame-env/bin/activate
```

Clone this repository to your machine. Then navigate to the following folder and install the dependencies:

```
$ git clone git@gitlab.uzh.ch:deborah.jakobi/linguistic-databases-project.git
$ cd linguistic-databases-project/lingu-wordnet-game
$ pip install .
```
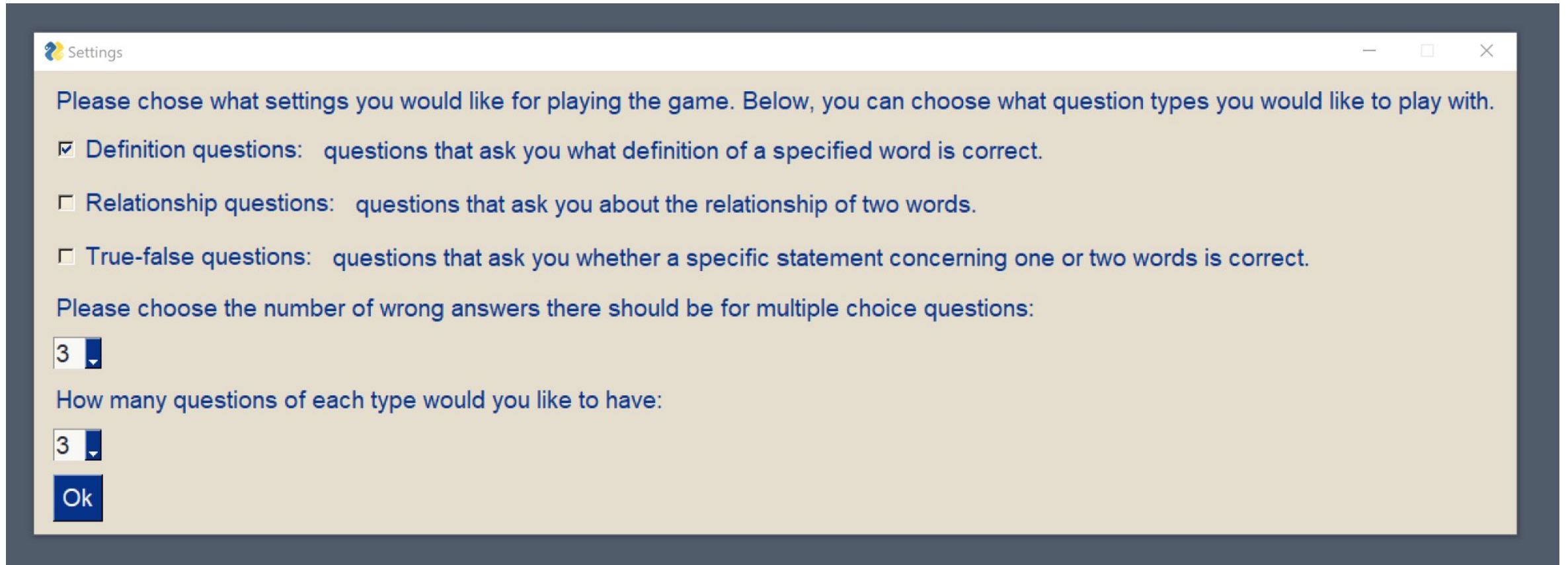
Before running the code, build the package:

```
$ python3 -m pip install --upgrade build
$ python3 -m build
```

Run the code like so:

```
$ python3 lingu-wordnet-game/lingu-wordnet-game.py
```

Backup demo

Settings screen

Multiple choice definition questions

Question 1

What is the correct definition of: accreditation

☑ 1) the act of granting credit or recognition (especially with respect to educational institution that maintains suitable standards)

☐ 2) the state of being enlarged

☐ 3) something you have to do

Show solution

Answer: correct answer

Solution 1

You selected answer 1)

Your answer is correct!

accreditation means: the act of granting credit or recognition (especially with respect to educational institution that maintains suitable standards)
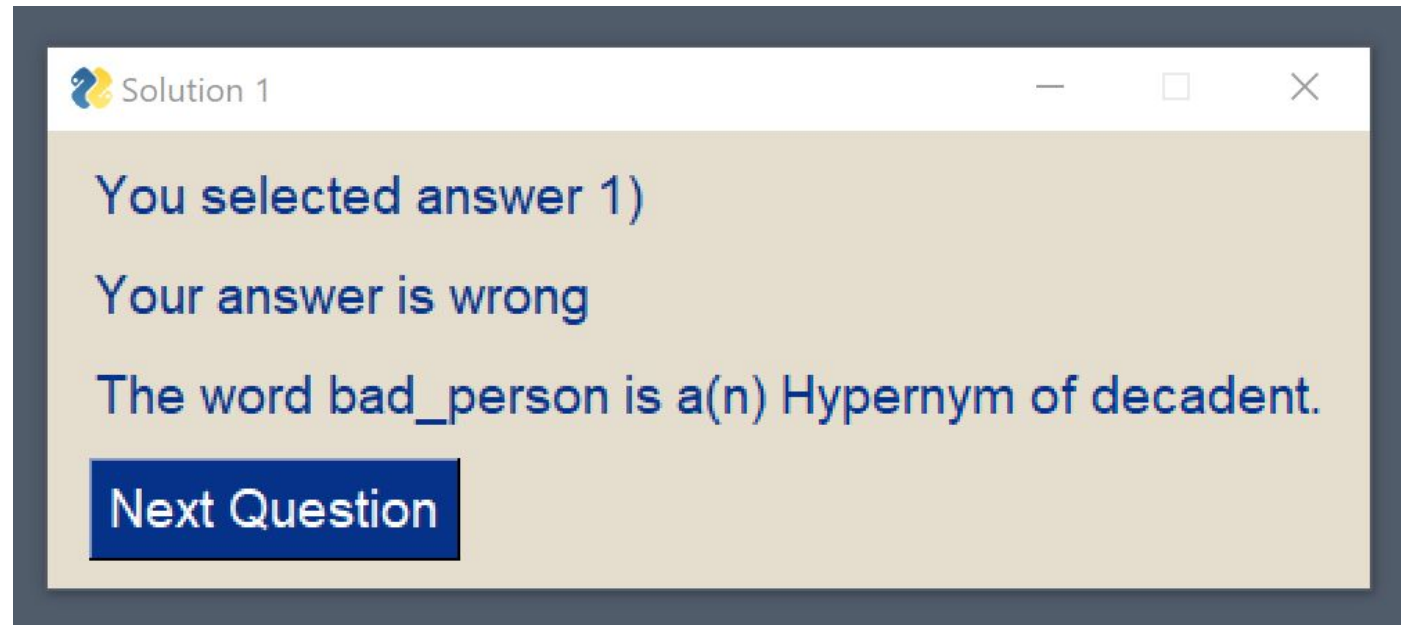
Next Question

Fill-in-the-blanks

The word 'bad_person' is a(n) _____ of 'decadent' (a person who has fallen into a decadent state (morally or artistically)).

- ☐ 1) Antonym

- ☐ 2) Synonym

- ☐ 3) Hypernym

Show solution

Answer: wrong answer
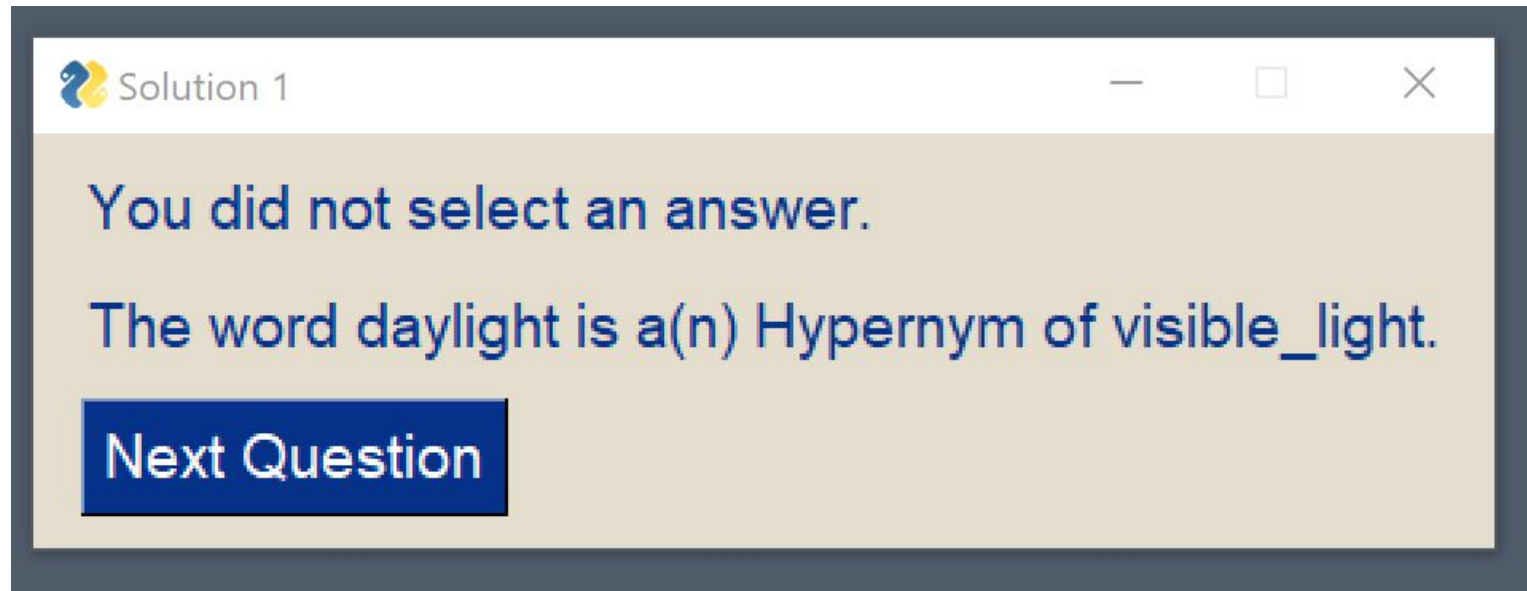
Relationship questions: true-false

The word 'daylight' is a(n) Hypernym of 'visible_light' (light during the daytime).

☐ 1) True

☐ 2) False

Show solution

Answer: no answer

Good-bye screen