# Creation of a Corpus of Specialised Language

Maria Luisa Dessì

# Corpus analysis

**QUANTITATIVE ANALYSIS**

## 1. What are the most significative items in that list?

### Frequency list of adjectives

| Qualifier | Frequency |
|---|---|
| linear | 78 |
| social | 76 |
| higher | 56 |
| violent | 50 |
| high | 45 |
| longitudinal | 43 |
| white | 42 |
| curvilinear | 39 |

| demonstrative | Frequency |
|---|---|
| 106 | these |
| 99 | this |

| possessive | frequency |
|---|---|
| their | 86 |

### Frequency list of adverbs

| adverbs | Frequency |
|---|---|
| as | 175 |
| more | 104 |
| not | 96 |
| early | 45 |
| however | 41 |
| only | 41 |

# Frequency list of conjunctions

| conjunctions | Frequency |
| --- | --- |
| and | 849 |
| or | 110 |
| also | 44 |
| but | 44 |

# Frequency list of nouns

| nouns | frequency |
| --- | --- |
| children | 312 |
| violence | 238 |
| aggression | 199 |
| tv | 150 |
| behavior | 108 |
| childhood | 93 |
| effects | 89 |
| levels | 88 |
| trajectories | 86 |
| child | 85 |
| teacher | 81 |
| age | 71 |
| adult | 70 |
| growth | 67 |
| intervention | 67 |
| exposure | 66 |
| development | 61 |
| school | 58 |
| risk | 57 |
| differences | 56 |
| significant | 50 |
| developmental | 48 |
| strategies | 47 |
| teachers | 47 |
| participants | 44 |
| classroom | 43 |
| year | 43 |
| three | 42 |

| | |
|---|---|
| models | 41 |
| boys | 40 |
| effect | 40 |
| media | 40 |
| ages | 39 |
| studies | 39 |
| instruction | 38 |

## Frequency list of prepositions

| prepositions | frequency |
|---|---|
| of | 952 |
| in | 618 |
| to | 539 |
| for | 347 |
| with | 195 |
| on | 183 |
| by | 139 |
| at | 103 |
| from | 93 |
| between | 86 |
| about | 40 |

## Frequency list of pronouns

| Demonstrative | frequency |
|---|---|
| that | 274 |
| **indefinite** | **frequency** |
| one | 44 |
| **personal** | **frequency** |
| they | 43 |
| it | 69 |
| we | 86 |
| **quantifier** | **frequency** |
| each | 41 |
| both | 90 |
| **relative** | **frequency** |

| | |
|---|---|
| which | 39 |
| who | 60 |

## Frequency list of verbs

| Verb | Frequency |
|---|---|
| is | 178 |
| were | 152 |
| are | 125 |
| be | 103 |
| viewing | 93 |
| was | 82 |
| have | 72 |
| change | 71 |
| study | 60 |
| reported | 58 |
| training | 52 |
| coaching | 50 |
| been | 45 |
| had | 44 |
| may | 41 |
| conduct | 40 |
| did | 38 |

## Frequency lists of articles

| Definite | Frequency |
|---|---|
| the | 1147 |
| **Indefinite** | **Frequency** |
| a | 308 |
| an | 54 |

**2. Create a core vocabulary of the target specialised language for each of the four word classes, i.e. nouns, verbs, adverbs and adjectives.**

**Core vocabulary:**

| Nouns | Adjectives | Verbs | Adverbs | Prepositions |
|---|---|---|---|---|
| children | aggressive | is | as | of |
| violence | these | were | more | in |
| aggression | this | are | not | to |
| tv | their | be | early | for |
| behavior | linear | viewing | however | with |
| childhood | social | was | only | on |
| effects | higher | have | | by |
| levels | violent | change | | at |
| trajectories | high | study | | from |
| child | longitudinal | reported | | between |
| teacher | white | training | | about |
| age | other | coaching | | |
| adult | curvilinear | been | | |
| growth | | had | | |
| intervention | | conduct | | |
| exposure | | did | | |
| development | | may | | |
| school | | | | |
| risk | | | | |
| differences | | | | |
| significant | | | | |
| developmental | | | | |
| strategies | | | | |
| teachers | | | | |
| participants | | | | |
| classroom | | | | |
| year | | | | |
| three | | | | |
| models | | | | |

| | | | | |
|---|---|---|---|---|
| boys | | | | |
| effect | | | | |
| media | | | | |
| ages | | | | |
| studies | | | | |
| instruction | | | | |

**3. The minimal core vocabulary, that is a limited number of items that are essential in the target variety of language, is the final outcome of the quantitative analysis carried on so far. It tells you what is used, but not how it is used.**

## QUALITATIVE ANALYSIS

**Focus on some of the most frequent words of your list, for example technical or semi-technical vocabulary, modals, verbs, connectors, etc.**

**Start observing their collocation.**

**Collocation of the first 10 nouns**

1. The Noun **children** is most of the times preceded by the adjectives: black white Hispanic, and followed by the noun exposure, by the relative who.., by the saxon genitive ex: children's development, children's growth, children's aggressive behavior ,etc.
2. The noun **violence** is often preceded by the noun exposure, media, childhood tv, adult tv, and followed by the word viewing and effects.
3. the noun **aggression** is often preceded by the preposition of, about, and the words adult, physical, female. It is followed by the conjunction and ex.: aggression and violence
4. The noun **tv** is used in this way: tv violence, tv habit, tv viewing, and Chilhood tv, adult tv, aggressive tv,

5. The noun **behavior** is preceded by the adjectives: prosocial, aggressive, violent, anti-social; it is followed by the verbs: revealed, reached, were characterized,

6. The noun **childhood** is used in this way: Childhood exposure, tv violence, aggression, media childhood, from childhood, during childhood, suggest that childhood, between chilhood.

7. the noun **effects** is used in this way: examining, predisposing, long-term, short-term effects; effects of children, classroom, exposure,etc.

8. the noun **levels** is used in this way: higher , lower equivalent levels; levels of competent, hostile, aggressive.

9. The noun **trajectories** is used in this way: developmental traj., children's traj., growth traj., traj. towards violence , of growth ., of aggressive ., of children, indicated that, etc.

10. the noun **child** is preceded by the prepositions: across, by, of, on, and followed by the words demographic, development, behavior, race.

## Collocation of the first 10 adjectives

1. The adjective **aggressive** is used in this way: measure of aggressive, the more aggressive.., their aggressive.., for aggressive.., development of aggressive, specific aggressive, future aggressive,

2. the possessive **these** is used in this way: nature of these.., change of these.., most of these.., each of these.., trajectories on these.., many of these..,

3. the possessive **this** is used in this way:for this article, of this study, in this study, in this domain, in this task,

4. The adjective **their** is used in this way: respect for their own culture, aware of their role, their exposure, of their role, to moderate their influence, their parents occupation, their parents aggressiveness,

5. The adjective **aggressive** is used in this way: aggrerssive fantasies, aggressive behavior, aggressive interpersonal negotiation, aggressive and violent behavior,

6. The adjective **linear** is used in this way: slower linear change, greater linear change, positive linear increase, negative linear increase, equivalent linear decline,

7. The adjective **social** is used in this way: children's social emotional development, positive social development, violence prevention and social emotional learning, social problem solving, social behavior, social value,

8. The adjective **higher** is used in this way: higher levels, significantly higher levels, followed by higher levels, evidencing higher levels,

9. The adjective **violent** is used in this way: violent behavior, aggressive and violent, the most violent, to use violent media, into a violent criminal, creating a violent environment, related to violent..,

10. The adjective **high** is used in this way: high lessons, high school, high tv violence viewers, who were high tv violence viewers, may score high.., in the high groups, in the high chilhood,

## Collocation of the first 10 verbs

The most frequent verbs used in this corpus are: **is**, **were**, **are**, **be**, **viewing**, **was**, **have**, **change**, **study**, **reported**. As we can see the verbs to be and to have are highly present. This suggests that in our syllabus we should focus our attention on the  correct use of these verbs. They are used most of the times in the passive form or in the present tenses.

### 1    Is there any linguistic pattern that is typical of the target language?

**Exposure** ( in everyday language we would say : children are exposed to tv violence or better we would say: children watch too many tv violent programs)

 "From prevention science, **knowledge has grown** about intervention strategies effective at reducing children's risk for future aggressive
    and violent behavior." (in everyday language we would say: we know more     about the risks of children's violent behavior)

### 5.  What are the most common forms of nominalization (if any)?

**Exposure**
**Measure**
**Improvement**
**Development**
**Assessment**
**Intervention**
**Evaluation**
**Prevention**

**Contribution**
**Negotiation**
**Reduction**
**Variation**
**Aggression**
**aggressiveness**

6. **From the observation of certain items, can you detect the presence of specific language functions essential in the target language?**

"**Child-reported  behavioral problems have also been used extensively in prior evaluations of preventive interventions,**" ( the use of passive form,)

 "**the perception of threat and experience of irritation or fear;   the accessibility of aggressive responses in one's memorial repertoire; and  the evaluation that aggression will lead to desirable positive consequences**" ( frequent  use of nominalizations)

We can also see the use of **modals** for example the use of **may** to express probability:
"other such measures **may** decrease
, and the precise nature of these changes
**may** vary**."**
"the assumption
of independence **may** have been violated"
"As such, there **may** be one or more
missing data points in any longitudinal design."

7. **Are there any passive forms? How often are they used and when?**

There are  many  **passive forms** such as the use of **perfect tenses** (**have been …ed  / has been …ed**)  and the **simple tenses** (**is/are …ed** or **was …ed /were ...ed** ]**.** They are used to explain the results or observations about studies or testing.

**Examples:**

➢ "Such mental processes **have been associated** with aggression and violence both currently"

➢ The RCCP, a universal, school-based intervention involving violence prevention and intergroup understanding, **was developed** collaboratively over time by Educators for Social Responsibility Metropolitan Area (ESR Metro) and the New York City Board of Education. Since its founding in, the RCCP **has served** over teachers and children in several hundred NewYork City public schools.

➢ "Many prevention programs **are** well **founded** on research-based theoretical principles, **are** rigorously **evaluated** for efficacy, and are often quite expensive,"

*8.* **What are the most common verbal tenses to be found?**

➢ **Passive forms**, used to explain the results or observations on studies or testing. The use of present tenses, some modals such as might, and may, the use of the future tenses.

9. **Can you detect compound noun phrases that are typical of this variety, (i.e. nominal phrases)?**

We can find, for example,

The growth trajectory for competent interpersonal negotiation strategies was marked by positive linear change.."
    "exposure to dramatic violence on TV.."
"extensive observation of violence around
    them.."

# *Conclusions*

My analysis of a corpus on psychology revealed the presence of many of the features of ESP such as the use of passive forms, nominalizations, present tenses, technical vocabulary, for this reason in planning a syllabus all these features must be taken into account.

Maria Luisa Dessì