# Search sheets: how to run the activity



The pages after this one are 120 **search sheets**, dealt from *three unnamed books* by three authors. Print them one-sided and hand out **one sheet per person**; every sheet is different, and nothing needs cutting out. Each holds 53–54 **token pairs**: a **next token** with the **previous token** that came before it in the text.

**Match the room to the sheets**, not the other way round. More people than sheets: print the whole set again, once per extra roomful. Every count multiplies equally, so the proportions — which is all the model is — hold. Fewer: hand the spares out anyway, two or three to a person. A sheet nobody holds is a hole in the model.

In market in, the boxed token is the context and *in* is the next token. A token keeps its colour wherever it appears, so say the colour as you call one out — "who has *in*? it's a grey one" — and the room can scan by colour before reading. With 8 colours and far more tokens, a colour narrows the search rather than settling it.

black grey red brown green blue purple magenta

## Running a round

Seed the board: someone reads out any token pair from their sheet, boxed token first, and both go up — the room supplies the beginning too. Then, each round:

1. **Call out the last token** on the board.
2. **Hands up** from everyone holding a matching token pair — one hand each.
3. **Pick a raised hand** — whoever shouts first is the fastest reader, not a sample of the text — and have that person choose one of their matching token pairs and read out its next token.
4. The scribe writes that token up. It is part of the next context. Repeat.

## What to watch for

**The hands approximate the probability distribution.** Every token pair sits with exactly one person, so a common continuation puts more hands in the air than a rare one; someone holding several matches still raises one hand, so it is close rather than exact. Nobody has to explain weighted sampling; the room is performing it.

**Nobody holds the model.** No single sheet can continue the text on its own. One hand is not a stall — but no hands at all means an empty seat, so reseed from any token pair and say why.

**The seam.** This corpus is three texts, tokenised apart so no token pair straddles the join. Nothing says which text a pair came from, and the model cannot tell either — so where two texts share a context, generation crosses between them unmarked.

6409 tokens, 1115 unique · 3.47 bits/token · perplexity 11.1 · branching factor 3.52

## Worked example

| | *the token pair that answers* | *on the board* |
|---|---|---|
| **Start anywhere.** Someone reads out a token pair — say this one — and both its tokens go on the board. | market in | market in |
| **Round 1.** Call out *in* — the last token on the board. A sheet across the room matches it, and that person reads out *the*. | in the | market in the |
| **Round 2.** Call out *the* — the last token on the board. A sheet across the room matches it, and that person reads out *bank*. | the bank | market in the bank |

Keep going for as long as you like. What the room writes will sound like *three unnamed books* without ever repeating it — that is the model generating, not the corpus being read out.